

Entropy production in full phase space for continuous stochastic dynamics

Richard E. Spinney and Ian J. Ford

*Department of Physics and Astronomy, UCL, Gower Street, London WC1E 6BT, UK and
London Centre for Nanotechnology, 17-19 Gordon Street London WC1H 0AH, UK*

(Dated: February 24, 2013)

The total entropy production and its three constituent components are described both as fluctuating trajectory-dependent quantities and as averaged contributions in the context of the continuous Markovian dynamics, described by stochastic differential equations with multiplicative noise, of systems with both odd and even coordinates with respect to time reversal, such as dynamics in full phase space. Two of these constituent quantities obey integral fluctuation theorems and are thus rigorously positive in the mean by Jensen's inequality. The third, however, is not and furthermore cannot be uniquely associated with irreversibility arising from relaxation, nor with the breakage of detailed balance brought about by non-equilibrium constraints. The properties of the various contributions to total entropy production are explored through the consideration of two examples: steady state heat conduction due to a temperature gradient, and transitions between stationary states of drift-diffusion on a ring, both in the context of the full phase space dynamics of a single Brownian particle.

PACS numbers: 05.70.Ln, 05.40.-a

I. INTRODUCTION

The concept of entropy was introduced over 150 years ago to provide a measure of the evident irreversibility of macroscopic thermodynamic phenomena. The conflict between its monotonic increase and the underlying time reversal symmetry of the microscopic dynamics, first pointed out by Loschmidt, is but one of its apparent mysteries. Nevertheless, in recent years significant insights into the nature of irreversibility and entropy production have emerged, partly due to the need for a framework to interpret thermodynamic processes on the nanoscale. These developments had their beginnings in the dissipation function and Fluctuation Theorem in deterministic thermostatted systems considered by Evans et al [1–4], and have continued with similar concepts within the realms of chaos theory [5] and of stochastic dynamical modelling [6, 7]. Some powerful results such as the Crooks and Jarzynski relations stand out [8–10] along with a unifying framework for overdamped Langevin dynamics [11] based on a stochastic description of the first law of thermodynamics commonly referred to as stochastic energetics [12]. In short, entropy production is a measure of the relative likelihoods of forward and reversed behaviour within the context of a model dynamical framework that includes specific dissipative terms. There are certain differences in viewpoint, but the central insight is that a mechanical (or maybe dynamical) quantity can be defined that matches the behaviour of thermodynamic entropy, in particular that on average, it increases with time. Its fluctuating nature provides additional insight into the behaviour of small systems.

More recently, it was proposed that the entropy production associated with non-equilibrium states of small systems, arising from an underlying stochastic

model of the dynamics, could be divided into two components, one related to relaxation (sometimes restricted to transitions between stationary states), and the other to any fundamental constraint that maintains the system away from an equilibrium [13–21]. The two components, termed adiabatic and non-adiabatic production rates respectively, were mapped onto earlier concepts known as excess and house-keeping heat transfers [19, 22]. In a recent development [23], however, it was shown that a third component of entropy production could be conceived, arising from the non-equilibrium constraint, but associated with relaxation towards the stationary state. It only arises when odd dynamical variables play a role in the dynamics, and even then only in specific cases. Only two of the three components of entropy production satisfy an integral fluctuation theorem (IFT), making them rigorously positive in the mean, properties shared by the sum of all three; the third, however, does not satisfy an IFT, and in the mean can take either sign.

In this paper, we develop these ideas further, within a framework of full phase space continuous dynamics modelled by stochastic differential equations, with the aim of pinning down the specific form of the three contributions, both in the mean and in fluctuations about the mean and go on to make use of the formalism in some instructive example systems.

II. THREE CONTRIBUTIONS TO ENTROPY PRODUCTION

We begin by considering the dynamics of a general set of variables $\mathbf{x} = (x_1, x_2, \dots, x_n)$ that may be odd or even under time reversal by considering the operation $\varepsilon\mathbf{x} = (\varepsilon_1 x_1, \varepsilon_2 x_2, \dots, \varepsilon_n x_n)$ where $\varepsilon_i = \pm 1$ for even and odd variables x_i respectively. Specifically, we consider

continuous Markovian dynamics described by a system of arbitrary uncorrelated Ito stochastic differential equations (SDEs) such that the evolution of the coordinates \mathbf{x} are given as

$$dx_i = A_i(\mathbf{x}, t)dt + B_i(\mathbf{x}, t)dW_i, \quad (1)$$

where dW_i denotes the Wiener process. Since we allow x_i to be either odd or even under time reversal we can divide the deterministic dynamics into reversible and irreversible components such that

$$dx_i = A_i^{\text{rev}}(\mathbf{x}, t)dt + A_i^{\text{ir}}(\mathbf{x}, t)dt + B_i(\mathbf{x}, t)dW_i \quad (2)$$

by defining

$$A_i^{\text{ir}}(\mathbf{x}, t) = \frac{1}{2} (A_i(\mathbf{x}, t) + \varepsilon_i A_i(\varepsilon\mathbf{x})) = \varepsilon_i A_i^{\text{ir}}(\varepsilon\mathbf{x}, t) \quad (3)$$

$$A_i^{\text{rev}}(\mathbf{x}, t) = \frac{1}{2} (A_i(\mathbf{x}, t) - \varepsilon_i A_i(\varepsilon\mathbf{x})) = -\varepsilon_i A_i^{\text{rev}}(\varepsilon\mathbf{x}, t). \quad (4)$$

We briefly note that we intend our notation, $A(\varepsilon\mathbf{x}, t)$, to imply a time reversal of all parameters that constitute A whether they be dynamical variables included in \mathbf{x} or not. For example, a term proportional to a magnetic field appearing in A_x , where x is an even spatial coordinate, would form part of A_x^{rev} since magnetic fields are odd with respect to time reversal in contrast with, for example, a force, F , which would appear in A_x^{ir} since force is even with respect to time reversal.

We have specified for simplicity that all our SDEs are driven by uncorrelated noise such that we have no cross derivatives in the corresponding Fokker-Planck equation. As such we may then represent the noise strengths as diffusion coefficients

$$\frac{1}{2}B_i(\mathbf{x}, t)^2 = D_i(\mathbf{x}, t) \quad (5)$$

that appear in a Fokker-Planck equation describing the joint probability density function of the coordinates

$$\begin{aligned} \frac{\partial p(\mathbf{x}, t)}{\partial t} &= \\ &- \sum_i \frac{\partial}{\partial x_i} (A_i(\mathbf{x}, t)p(\mathbf{x}, t)) + \sum_i \frac{\partial^2}{\partial x_i^2} (D_i(\mathbf{x}, t)p(\mathbf{x}, t)). \end{aligned} \quad (6)$$

For simplicity in the later development we assume that the diffusion coefficient is symmetric with respect to time reversal such that $D(\varepsilon\mathbf{x}) = D(\mathbf{x})$, which puts no restriction on the dependence on even coordinates, but requires that $D(\mathbf{x})$ is an even function of any odd coordinates.

It is helpful to express the Fokker-Planck equation

as a continuity equation in terms of the probability density current $J(\mathbf{x}, t)$

$$\begin{aligned} \frac{\partial p(\mathbf{x}, t)}{\partial t} &= -\nabla \cdot J(\mathbf{x}, t) \\ &= -\nabla \cdot (J^{\text{ir}}(\mathbf{x}, t) + J^{\text{rev}}(\mathbf{x}, t)) \end{aligned} \quad (7)$$

which we separate into irreversible and reversible components. These take vector form $J = (J_1, J_2, \dots, J_n)$ as do the drift and diffusion coefficients $A = (A_1, A_2, \dots, A_n)$ and $D = (D_1, D_2, \dots, D_n)$ such that

$$\begin{aligned} J^{\text{ir}}(\mathbf{x}, t) &= A^{\text{ir}}(\mathbf{x}, t)p(\mathbf{x}, t) - \nabla \cdot (D(\mathbf{x}, t)p(\mathbf{x}, t)) \\ J^{\text{rev}}(\mathbf{x}, t) &= A^{\text{rev}}(\mathbf{x}, t)p(\mathbf{x}, t). \end{aligned} \quad (8)$$

Having set out the dynamics we shall be using, we now consider the general procedure for producing quantities which obey IFTs. Given an interval of duration τ , such a quantity consists of a difference between the logarithmic probability density of a given trajectory under what we shall term the forward dynamics, time dependence of the dynamics (equivalent here to an external protocol) and initial distribution of starting configurations, and that of another appropriately chosen trajectory under suitable dynamics, protocol and initial distribution. We write $p[\vec{x}]$ as the probability density of the forward trajectory or path, $\vec{x} = \mathbf{x}(t)$ for $0 \leq t \leq \tau$, with a probability density function of starting configurations, $p(\mathbf{x}(0), 0)$, which acts as an initial condition for the Fokker-Planck equation introduced earlier. A quantity that obeys an IFT is then of the form

$$A[\vec{x}] = \ln [p[\vec{x}] / p^*[\vec{x}^*]] \quad (9)$$

where $p^*[\vec{x}^*]$ is the probability density of a path, \vec{x}^* , under chosen dynamics (with specified nature and time dependence) and initial condition. Demonstrating the adherence of such a quantity to an IFT with respect to the forward dynamics and time dependence is straightforward by the reasoning

$$\begin{aligned} \langle \exp [-A[\vec{x}]] \rangle &= \int d\vec{x} p[\vec{x}] \exp [-A[\vec{x}]] \\ &= \int d\vec{x} p[\vec{x}] \frac{p^*[\vec{x}^*]}{p[\vec{x}]} \\ &= \int d\vec{x}^* p^*[\vec{x}^*] = 1. \end{aligned} \quad (10)$$

Such a result requires a Jacobian of unity for the path transformation $\vec{x} \rightarrow \vec{x}^*$ in order for the path integrals to be equivalent (a result assured for any involutive transformation) and the requirement $p^*[\vec{x}^*] = 0$ for all $p[\vec{x}] = 0$ ensuring that for normalised p and p^* all possible paths under the dynamics that produce p^* are contained within the final integral. This may be seen as a more general version of the ergodic consistency requirement [24] since it is a condition required for the appropriate inclusion of paths in the final integral of Eq. (10), but we do not demand that p and p^* be

non-zero for all paths. Further, any quantity $A[\vec{x}]$ based on a transformation with a Jacobian of unity takes the same form irrespective of whether probabilities or probability densities are used in the construction owing to the equivalence of measure. Such a quantity is therefore constructed unambiguously and is the direct analogue of the same quantity defined in discrete space [19, 23]. The implication of the positivity in the mean, $\langle A[\vec{x}] \rangle \geq 0$, of the quantity $A[\vec{x}]$ is assured by Jensen's inequality. We point out that the path, \vec{x}^* , and the dynamics must be carefully chosen in order to satisfy the requirements and to produce a physically meaningful quantity. This is particularly relevant in the presence of both odd and even variables as for many physical systems the common choice of reverse path $\vec{x}^*(t) = \vec{x}(\tau-t)$ cannot be generated under the forward dynamics, and if used to define a quantity of the form in Eq. (9) would render the final integral in Eq. (10) equal to zero. However, the choice $\vec{x}^*(t) = \varepsilon\vec{x}(\tau-t)$ typically can be generated and so leads to an IFT. However, this choice of the second path, $\vec{x}^*(t) = \varepsilon\vec{x}(\tau-t)$, is appropriate not just because of the guarantee of an IFT, but because it means the constructed quantity $A[\vec{x}]$ serves as a measure of the irreversibility of the process and thus characterises the total entropy production.

By following the above rules, and making the definitions $\vec{x}^\dagger(t) = \varepsilon\vec{x}(\tau-t)$, $\vec{x}^R(t) = \vec{x}(\tau-t)$ and $\vec{x}^T(t) = \varepsilon\vec{x}(t)$, we may construct dimensionless entropy changes, which are thermodynamically meaningful when multiplied by k_B , of the form in Eq. (9):

$$\begin{aligned}\Delta S_{\text{tot}} &= \ln p[\vec{x}] - \ln p^R[\vec{x}^\dagger] \\ &= \ln \frac{p(\vec{x}(0), 0)}{p(\vec{x}(\tau), \tau)} + \ln \frac{p[\vec{x}(\tau)|\vec{x}(0)]}{p^R[\varepsilon\vec{x}(0)|\varepsilon\vec{x}(\tau)]}\end{aligned}\quad (11)$$

$$\begin{aligned}\Delta S_1 &= \ln p[\vec{x}] - \ln p^{\text{ad},R}[\vec{x}^R] \\ &= \ln \frac{p(\vec{x}(0), 0)}{p(\vec{x}(\tau), \tau)} + \ln \frac{p[\vec{x}(\tau)|\vec{x}(0)]}{p^{\text{ad},R}[\vec{x}(0)|\vec{x}(\tau)]}\end{aligned}\quad (12)$$

$$\begin{aligned}\Delta S_2 &= \ln p[\vec{x}] - \ln p^{\text{ad}}[\vec{x}^T] \\ &= \ln \frac{p(\vec{x}(0), 0)}{p(\vec{x}(0), 0)} + \ln \frac{p[\vec{x}(\tau)|\vec{x}(0)]}{p^{\text{ad}}[\varepsilon\vec{x}(\tau)|\varepsilon\vec{x}(0)]}\end{aligned}\quad (13)$$

such that the total path probability densities are divided into initial probability density distributions and conditional probability densities. In the above the label R designates a reversed protocol, equivalent here to reversed time dependence in the dynamics, and ‘ad’ designates that the dynamics are so-called adjoint with respect to the forward dynamics, defined as the dynamics which reach the same stationary state, but with the opposite stationary current [15, 19, 24]. All are expected to obey IFTs by the nature of their form. Then, by the construc-

tion $\Delta S_{\text{tot}} = \Delta S_1 + \Delta S_2 + \Delta S_3$, we define

$$\begin{aligned}\Delta S_3 &= \ln p^{\text{ad}}[\vec{x}^T] + \ln p^{\text{ad},R}[\vec{x}^R] - \ln p[\vec{x}] - \ln p^R[\vec{x}^\dagger] \\ &= \ln \frac{p^{\text{ad},R}[\vec{x}(0)|\vec{x}(\tau)]p^{\text{ad}}[\varepsilon\vec{x}(\tau)|\varepsilon\vec{x}(0)]}{p[\vec{x}(\tau)|\vec{x}(0)]p^R[\varepsilon\vec{x}(0)|\varepsilon\vec{x}(\tau)]}\end{aligned}\quad (14)$$

which cannot be expressed in the form of Eq. (9) and so does not obey an IFT. By following the formalism of Seifert [11, 25] we identify

$$\begin{aligned}\Delta S_{\text{tot}} &= \ln \frac{p(\vec{x}(0), 0)}{p(\vec{x}(\tau), \tau)} + \int_{t=0}^{t=\tau} d \left(\frac{\Delta Q}{k_B T_{\text{env}}(\vec{x}(t), t)} \right) \\ &= \Delta S_{\text{sys}} + \Delta S_{\text{med}},\end{aligned}\quad (15)$$

where ΔS_{sys} is known as the change in dimensionless system entropy and ΔS_{med} is a generalisation of the dimensionless entropy production in the environment, or medium, for an environmental temperature, T_{env} , which we allow to be phase space or time dependent (a property we employ in example I). The integral should be interpreted as a total medium entropy change, considered as the sum of all the incremental heat transfers to separate fixed temperature heat baths to which the particle is exposed over the course of its trajectory, divided by the appropriate temperature. Of course, this reproduces the usual $\Delta S_{\text{med}} = \Delta Q/k_B T_{\text{env}}$ of stochastic thermodynamics [11] when the temperature is constant. Furthermore, to make connection with a key concept in non-equilibrium thermodynamics, we may divide the heat transfer to the environment (considering one at a constant temperature for clarity without loss of generality), into the so-called excess and house-keeping heats according to the formalism of Oono and Paniconi [22] such that for a given environmental temperature $\Delta Q = \Delta Q_{\text{ex}} + \Delta Q_{\text{hk}}$. To align our quantities with such a formalism we associate ΔS_1 with the excess heat

$$\Delta Q_{\text{ex}} = (\Delta S_1 - \Delta S_{\text{sys}})k_B T_{\text{env}},\quad (16)$$

ΔS_2 with a so called ‘generalised house-keeping heat’ [23]

$$\Delta Q_{\text{hk,G}} = \Delta S_2 k_B T_{\text{env}}\quad (17)$$

and ΔS_3 with the ‘transient house-keeping heat’,

$$\Delta Q_{\text{hk,T}} = \Delta S_3 k_B T_{\text{env}},\quad (18)$$

named to reflect its mean behaviour, such that $\Delta Q_{\text{hk}} = \Delta Q_{\text{hk,G}} + \Delta Q_{\text{hk,T}}$. Like previous formalisms [19, 21] where the entropy production was divided into two contributions associated with relaxation, and an absence of detailed balance, respectively, both in the mean and in detail [19–21], we have a contribution ΔS_1 which is non-zero only in the presence of relaxation and ΔS_2 which is non-zero only in the absence of detailed balance both in the mean and in detail. However, we also have a quantity ΔS_3 which is non-zero in detail only in the absence of detailed balance, but only contributes in the mean during the course of relaxation. Such a formalism asserts that

the two origins of entropy production may often be more closely related, with such a circumstance arising under the inclusion of odd variables and when the stationary distribution is asymmetric in any of those odd variables. The aim of this paper is to derive the equations of motion for each of these quantities for continuous stochastic systems and to illustrate their behaviour through some simple examples.

III. REPRESENTING ENTROPY PRODUCTION FOR CONTINUOUS BEHAVIOUR

Since we are describing the dynamics using SDEs it is sensible to seek a description of a small increment in each entropy production given an increment in the underlying variables $\mathbf{x}' - \mathbf{x} = \mathbf{x}(t+dt) - \mathbf{x}(t)$ in a time dt so that we identify from Eqs. (11), (12), (13) and (14)

$$d\Delta S_{\text{tot}} = -d(\ln p) + \ln \frac{p(\mathbf{x}', t+dt|\mathbf{x}, t)}{p(\varepsilon\mathbf{x}, t+dt|\varepsilon\mathbf{x}', t)} \quad (19)$$

$$d\Delta S_1 = -d(\ln p) + \ln \frac{p(\mathbf{x}', t+dt|\mathbf{x}, t)}{p^{\text{ad}}(\mathbf{x}, t+dt|\mathbf{x}', t)} \quad (20)$$

$$d\Delta S_2 = \ln \frac{p(\mathbf{x}', t+dt|\mathbf{x}, t)}{p^{\text{ad}}(\varepsilon\mathbf{x}', t+dt|\varepsilon\mathbf{x}, t)} \quad (21)$$

$$d\Delta S_3 = \ln \frac{p^{\text{ad}}(\mathbf{x}, t+dt|\mathbf{x}', t)p^{\text{ad}}(\varepsilon\mathbf{x}', t+dt|\varepsilon\mathbf{x}, t)}{p(\mathbf{x}', t+dt|\mathbf{x}, t)p(\varepsilon\mathbf{x}, t+dt|\varepsilon\mathbf{x}', t)} \quad (22)$$

thereby establishing the SDEs that describe entropy production and noting the abbreviation $d(\ln(p)) = \ln(p(\mathbf{x}(t+dt), t+dt)/p(\mathbf{x}(t), t))$.

To proceed we require a representation of the path probabilities in these expressions valid over the small time interval dt . This may be achieved by considered the short time Green's function or 'short time propagator' [26] which is given generally as the conditional probability of a displacement $dx = x' - x$ in a time dt subject to a delta function initial condition and is of the form

$$p(\mathbf{x}', t+dt|\mathbf{x}, t) = \prod_i \sqrt{\frac{1}{4\pi D_i(\mathbf{r}, t)dt}} \times \exp \left[\frac{(dx_i - A_i(\mathbf{r}, t)dt + 2a(\partial D_i(\mathbf{r}, t)/\partial r_i)dt)^2}{4D_i(\mathbf{r})dt} - adt \frac{\partial A_i(\mathbf{r}, t)}{\partial r_i} + a^2 dt \frac{\partial^2 D_i(\mathbf{r}, t)}{\partial r_i^2} \right] \quad (23)$$

where $dx_i = x'_i - x_i$ and where a is a free parameter ranging from 0 to 1 which defines the evaluation point of certain terms in the propagator $\mathbf{r} = a\mathbf{x}' + (1-a)\mathbf{x}$ and $r_i = ax'_i + (1-a)x_i$, and which reflects the ambiguity of a discretised interpretation of continuous stochastic behaviour. We note, however that as $dt \rightarrow 0$ all forms for the propagator are correct: they are all accurate to first order in dt and result in the same Fokker-Planck equation.

We wish to construct the increment in entropy production in the medium by a consideration of

$$d\Delta S_{\text{med}} = \ln \frac{p(\mathbf{x}', t+dt|\mathbf{x}, t)}{p(\varepsilon\mathbf{x}, t+dt|\varepsilon\mathbf{x}', t)} \quad (24)$$

by employing the appropriate reverse short time propagator

$$p(\varepsilon\mathbf{x}, t+dt|\varepsilon\mathbf{x}', t) = \prod_i \sqrt{\frac{1}{4\pi D_i(\varepsilon\mathbf{r}', t)dt}} \exp \left[-\frac{(-\varepsilon_i dx_i - A_i(\varepsilon\mathbf{r}', t)dt + 2b(\partial D_i(\varepsilon\mathbf{r}', t)/\partial(\varepsilon_i r'_i))dt)^2}{4D_i(\varepsilon\mathbf{r}')dt} - bdt \frac{\partial A_i(\varepsilon\mathbf{r}', t)}{\partial(\varepsilon_i r'_i)} + b^2 dt \frac{\partial^2 D_i(\varepsilon\mathbf{r}', t)}{\partial(\varepsilon_i r'_i)^2} \right] \quad (25)$$

where b is a corresponding free parameter ranging from 0 to 1 such that $\mathbf{r}' = b\mathbf{x} + (1-b)\mathbf{x}'$ and $r'_i = bx_i + (1-b)x'_i$. Using Eqs. (3) and (4) along with the assumption $D_i(\varepsilon\mathbf{x}) = D_i(\mathbf{x})$ such that any diffusion constants for odd variables are symmetric, we may write

$$p(\varepsilon\mathbf{x}, t+dt|\varepsilon\mathbf{x}', t) = \prod_i \sqrt{\frac{1}{4\pi D_i(\mathbf{r}', t)dt}} \exp \left[-\frac{(-\varepsilon_i dx_i - \varepsilon_i (-A_i^{\text{rev}}(\mathbf{r}', t) + A_i^{\text{ir}}(\mathbf{r}', t))dt + 2b(\partial D_i(\mathbf{r}', t)/\partial(\varepsilon_i r'_i))dt)^2}{4D_i(\mathbf{r}')dt} - bdt \left(\frac{\partial \varepsilon_i A_i^{\text{ir}}(\mathbf{r}', t)}{\partial(\varepsilon_i r'_i)} - \frac{\partial \varepsilon_i A_i^{\text{rev}}(\mathbf{r}', t)}{\partial(\varepsilon_i r'_i)} \right) + b^2 dt \frac{\partial^2 D_i(\mathbf{r}', t)}{\partial(\varepsilon_i r'_i)^2} \right] \quad (26)$$

which is the same as

$$p(\varepsilon \mathbf{x}, t + dt | \varepsilon \mathbf{x}', t) = \prod_i \sqrt{\frac{1}{4\pi D_i(\mathbf{r}', t)dt}} \exp \left[-\frac{(-dx_i - (-A_i^{\text{rev}}(\mathbf{r}', t) + A_i^{\text{ir}}(\mathbf{r}', t))dt + 2b(\partial D_i(\mathbf{r}', t)/\partial r'_i)dt)^2}{4D_i(\mathbf{r}')dt} \right. \\ \left. - bdt \left(\frac{\partial A_i^{\text{ir}}(\mathbf{r}', t)}{\partial r'_i} - \frac{\partial A_i^{\text{rev}}(\mathbf{r}', t)}{\partial r'_i} \right) + b^2 dt \frac{\partial^2 D_i(\mathbf{r}', t)}{\partial r'^2_i} \right]. \quad (27)$$

The mathematical details necessary for the development of Eq. (24), which due to their somewhat cumbersome nature we leave to appendix A, reveal that for multiplicative noise one obtains a result which is dependent on the choice a and b . The resolution of this apparent arbitrariness is not related to the nature of the underlying SDEs, but rather on consistently using the equivalent evaluation point for forward and time reversed paths on an infinitesimal scale. The normal rules of calculus would dictate no dependence, but different rules apply to SDEs and stochastic calculus. Such a consideration reveals the correct choice in Eq. (27) to be $b = 1 - a$ with a remaining as a free parameter. This yields the unambiguous Ito SDE for the medium entropy change

$$d\Delta S_{\text{med}} = \sum_i \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dx_i - \frac{A_i^{\text{rev}}(\mathbf{x}) A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dt \\ + \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} dt - \frac{\partial A_i^{\text{rev}}(\mathbf{x})}{\partial x_i} dt \\ - \frac{1}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dx_i \\ + \frac{(A^{\text{rev}}(\mathbf{x}) - A^{\text{ir}}(\mathbf{x}))}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dt \\ - \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} dt + \frac{1}{D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 dt \quad (28)$$

where for brevity we use notation $f(\mathbf{x}) \equiv f(\mathbf{x}, t)$. To clarify, in such an approach choices may include Stratonovich ($a = b = 1/2$) evaluation for both propagators in Eq. (24), a choice which is implicitly used by many authors [15, 25] within integrated Onsager-Machlup approaches, but does not preclude others in the construction of SDEs such as, for example, an Ito prescription ($a = 0$) in the forward propagator and a Hanggi-Klimontovich ($b = 1$) in the backwards propagator. We point out that all evaluation points lead to the correct *path probability* when supplemented with the correct multiplication scheme, but that if one has multiplicative noise, the correct representation of the *entropy production* requires the more exact relation between the evaluation points.

Proceeding, we may now construct an SDE for the total entropy production by first considering an increment in the system entropy which under Ito rules

is

$$d\Delta S_{\text{sys}} = -d(\ln p(\mathbf{x})) \\ = -\frac{1}{p(\mathbf{x})} \frac{\partial p(\mathbf{x})}{\partial t} dt - \frac{1}{p(\mathbf{x})} \sum_i \frac{\partial p(\mathbf{x})}{\partial x_i} dx_i \\ - \sum_i \frac{D_i(\mathbf{x})}{p(\mathbf{x})} \left(\frac{\partial^2 p(\mathbf{x})}{\partial x_i^2} - \frac{1}{p(\mathbf{x})} \left(\frac{\partial p(\mathbf{x})}{\partial x_i} \right)^2 \right) dt, \quad (29)$$

which together with Eq. (28), and after insertion of the Fokker-Planck equation, leads to

$$d\Delta S_{\text{tot}} = \sum_i -\frac{1}{p(\mathbf{x})} \frac{\partial p(\mathbf{x})}{\partial x_i} dx_i + \frac{1}{p(\mathbf{x})} \frac{\partial(A_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i} dt \\ - \frac{1}{p(\mathbf{x})} \left(\frac{\partial^2(D_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i^2} + D_i(\mathbf{x}) \frac{\partial^2 p(\mathbf{x})}{\partial x_i^2} \right. \\ \left. - \frac{D_i(\mathbf{x})}{p(\mathbf{x})} \left(\frac{\partial p(\mathbf{x})}{\partial x_i} \right)^2 \right) dt \\ + \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dx_i - \frac{A_i^{\text{rev}}(\mathbf{x}) A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dt \\ + \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} dt - \frac{\partial A_i^{\text{rev}}(\mathbf{x})}{\partial x_i} dt \\ - \frac{1}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dx_i \\ + \frac{(A^{\text{rev}}(\mathbf{x}) - A^{\text{ir}}(\mathbf{x}))}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dt \\ - \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} dt + \frac{1}{D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 dt. \quad (30)$$

If Stratonovich rules, for example, are preferred we can write (by definition of the Stratonovich integral, indicated by the \circ notation)

$$d\Delta S_{\text{tot}} = \sum_i -\frac{1}{p(\mathbf{x})} \frac{\partial p(\mathbf{x})}{\partial x_i} \circ dx_i + \frac{1}{p(\mathbf{x})} \frac{\partial(A_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i} dt \\ - \frac{1}{p(\mathbf{x})} \left(\frac{\partial^2(D_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i^2} \right) dt \\ + \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} \circ dx_i - \frac{A_i^{\text{rev}}(\mathbf{x}) A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dt \\ - D_i(\mathbf{x}) \frac{\partial}{\partial x_i} \left(\frac{A_i^{\text{rev}}(\mathbf{x})}{D_i(\mathbf{x})} \right) dt \\ - \frac{1}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} \circ dx_i. \quad (31)$$

This is a very general and robust definition of the entropy production for continuous stochastic behaviour and can be thought of as a generalisation of the pioneering approach in [11] wherein the equation of motion for entropy essentially describes $d\Delta S_{\text{tot}}$ for a specific system with additive noise, even variables ($\varepsilon \mathbf{x} = \mathbf{x}$) and implicitly using Stratonovich rules.

We point out that such a construction allows us to consider purely deterministic coordinates ($D_i(\mathbf{x}) = 0$) as would apply, for example, to the case of spatial coordinates within a full phase space Langevin description. In such coordinates $D_i(\mathbf{x})$ is assumed constant and taken to zero. The remaining terms then clearly diverge unless we demand $A_i^{\text{ir}}(\mathbf{x}) = 0$ since in these instances, for the reverse path to be a solution to the forward dynamics the motion must be purely reversible. This condition simply amounts to the requirement that the reverse path exists. There is, however, a contribution to the medium entropy production, due to the dynamics of these coordinates, technically since path probability densities, not probabilities, are being considered in the formulation. The contribution to the medium entropy production due to the deterministic behaviour of these coordinates is

$$\Delta S_{\text{med,det}} = -\frac{\partial A_i^{\text{rev}}(\mathbf{x})}{\partial x_i} dt, \quad (32)$$

a result that provides an insight into the similarities and differences between stochastic and deterministic measures of irreversibility: it is demonstrably equal to the phase space contraction found in non-linear dynamical systems, which is associated with the heat transfer to the environment brought about by thermostating terms in such approaches. This leads to a quantity that is positive in the mean for deterministic systems: the dissipation function [2]. We point out, however that *total entropy production*, as defined here for stochastic systems, is zero for deterministic dynamics. This is because the change in the system entropy would be equal and opposite to the change in medium entropy, technically since it involves probability densities at the start and end of the process. In contrast the dissipation function can provide a measure of irreversibility because it involves a comparison of trajectories originating *from the same* starting distribution. This contrast is to be expected as the total entropy production, as defined for the systems we consider, arises from explicit irreversibility in the dynamics, which deterministic, reversible equations do not provide.

Frequently the average entropy production rate is argued to be proportional to the mean probability flux squared, as derived, for example, by taking the time derivative of the Gibbs entropy of a system, and identifying an evidently positive contribution as the total entropy production rate and the remainder as the (negative of) the medium entropy production rate

[21, 27]. We prefer however, to derive the average contributions directly from the SDEs so that we can avoid arbitrarily identifying a positive contribution with a quantity expected to obey an IFT: strictly there is no guarantee such a division is unique, as another description shows [28]. To do so is straightforward and requires us to find the average increment in ΔS_{tot} by means of the integral

$$\langle d\Delta S_{\text{tot}} \rangle = \int d\mathbf{x} \int d\mathbf{x}' p(\mathbf{x}, t) p(\mathbf{x}', t + dt | \mathbf{x}, t) d\Delta S_{\text{tot}}. \quad (33)$$

The benefit of such a formulation is that we may characterise $d\Delta S_{\text{tot}}$ using an Ito SDE based on the underlying relations $dx_i = A_i dt + B_i dW_i$ and then use the martingale property of the Ito stochastic integral $\langle B_i dW_i \rangle = 0$ since B_i is non-anticipating, such that we can simplify the integral in Eq. (33) by writing

$$\langle d\Delta S_{\text{tot}} \rangle = \int d\mathbf{x} p(\mathbf{x}) \langle d\Delta S_{\text{tot}} | \mathbf{x} \rangle \quad (34)$$

and evaluating the conditional average $\langle d\Delta S_{\text{tot}} | \mathbf{x} \rangle$ by replacing all occurrences of dx_i with $(A_i^{\text{ir}} + A_i^{\text{rev}})dt$ in $d\Delta S_{\text{tot}}$. We thus get

$$\begin{aligned} \langle d\Delta S_{\text{tot}} \rangle = & \sum_i \left[\int d\mathbf{x} \frac{p(\mathbf{x})(A_i^{\text{ir}}(\mathbf{x}))^2}{D_i(\mathbf{x})} + 2p(\mathbf{x}) \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} \right. \\ & - 2p(\mathbf{x}) \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} - p(\mathbf{x}) \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} \\ & + \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 - \frac{\partial^2 D_i(\mathbf{x}) p(\mathbf{x})}{\partial x_i^2} \\ & \left. - D_i(\mathbf{x}) \frac{\partial^2 p(\mathbf{x})}{\partial x_i^2} + \frac{D_i(\mathbf{x})}{p(\mathbf{x})} \left(\frac{\partial p(\mathbf{x})}{\partial x_i} \right)^2 \right] dt. \end{aligned} \quad (35)$$

By applying the product rule, integrating by parts and assuming $p(\mathbf{x})$ and $\partial p(\mathbf{x})/\partial x_i$ either vanish or cancel at the boundaries, we may simplify to find the total entropy production rate

$$\begin{aligned} \frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = & \sum_i \int d\mathbf{x} \frac{\left(p(\mathbf{x}) A_i^{\text{ir}}(\mathbf{x}) - D_i(\mathbf{x}) \frac{\partial p(\mathbf{x})}{\partial x_i} - p(\mathbf{x}) \frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2}{p(\mathbf{x}) D_i(\mathbf{x})} \end{aligned} \quad (36)$$

or more concisely

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = \sum_i \int d\mathbf{x} \frac{(J_i^{\text{ir}}(\mathbf{x}))^2}{p(\mathbf{x}) D_i(\mathbf{x})}, \quad (37)$$

providing an expression for the mean instantaneous entropy production rate which is rigorously positive, as it must be because of the adherence of ΔS_{tot} to an IFT, and is dependent on the *irreversible* flux.

IV. EXPRESSIONS FOR ΔS_1 , ΔS_2 AND ΔS_3

In order to consider a division of the entropy production into the thermodynamically meaningful quantities outlined above we are required to construct path probabilities using the so-called adjoint dynamics. These dynamics may not be physically realisable: for example they may require negative positional steps to result from positive velocities (as indicated by the paths $\mathbf{x}^R(t)$ and $\mathbf{x}^T(t)$), but this is of no concern since they are only introduced for the mathematical construction of the entropy contributions. We consider an arbitrary stationary distribution of a given system which may be written in terms of a non-equilibrium potential, $\phi(\mathbf{x})$, such that

$$p^{\text{st}}(\mathbf{x}) = \exp[-\phi(\mathbf{x})] \quad (38)$$

and assert that the adjoint dynamics are those that result in the same stationary distribution, but have an opposite flux. As such we require

$$\frac{\partial p^{\text{st}}(\mathbf{x})}{\partial t} = -\nabla \cdot J^{\text{st}}(\mathbf{x}) = \nabla \cdot J^{\text{st,ad}}(\mathbf{x}) = 0 \quad (39)$$

with

$$J^{\text{st,ad}}(\mathbf{x}) = -J^{\text{st}}(\mathbf{x}). \quad (40)$$

In order to characterise the adjoint dynamics we construct the adjoint flux according to

$$\begin{aligned} J^{\text{st,ad}}(\mathbf{x}) &= A^{\text{ad}}(\mathbf{x})p^{\text{st}}(\mathbf{x}) - \frac{\partial}{\partial x_i}(D_i(\mathbf{x})p^{\text{st}}(\mathbf{x})) \\ &= A^{\text{ad}}(\mathbf{x})e^{-\phi(\mathbf{x})} - \frac{\partial}{\partial x_i}(D_i(\mathbf{x})e^{-\phi(\mathbf{x})}) \\ &= \left(A^{\text{ad}}(\mathbf{x}) - \frac{\partial D_i(\mathbf{x})}{\partial x_i} + D_i(\mathbf{x})\frac{\partial \phi(\mathbf{x})}{\partial x_i} \right) e^{-\phi(\mathbf{x})} \\ &= -\left(A(\mathbf{x}) - \frac{\partial D_i(\mathbf{x})}{\partial x_i} + D_i(\mathbf{x})\frac{\partial \phi(\mathbf{x})}{\partial x_i} \right) e^{-\phi(\mathbf{x})}. \end{aligned} \quad (41)$$

Consequently we have the requirement

$$A^{\text{ad}}(\mathbf{x}) = -A(\mathbf{x}) + 2\frac{\partial D_i(\mathbf{x})}{\partial x_i} - 2D_i(\mathbf{x})\frac{\partial \phi(\mathbf{x})}{\partial x_i}. \quad (42)$$

Let us now consider the quantity

$$d\Delta S_{\text{ex}} = \ln \frac{p(\mathbf{x}', t + dt | \mathbf{x}, t)}{p^{\text{ad}}(\mathbf{x}, t + dt | \mathbf{x}', t)}, \quad (43)$$

where $\Delta S_{\text{ex}} = \Delta Q_{\text{ex}}/k_B T_{\text{env}}$, which we have previously asserted constitutes part of the incremental contribution to the quantity ΔS_1 based on relations in Eqs. (12) and its short time representation. We evaluate Eq. (43), taking the transition probability density in the numerator from Eq. (23) and, for convenience, choosing $a = 1/2$. We can represent the transition probability density appearing in the denominator through a similar construction, but using a substitution for the adjoint drift term from Eq. (42), together with the complementary evaluation point choice $b = 1 - a = 1/2$ such that

$$\begin{aligned} p^{\text{ad}}(\mathbf{x}, t + dt | \mathbf{x}', t) &= \prod_i \sqrt{\frac{1}{4\pi D_i(\mathbf{r})dt}} \exp \left[-\frac{(-dx_i + (A_i(\mathbf{r}) - 2(\partial D_i(\mathbf{r})/\partial r_i) + 2D_i(\mathbf{r})(\partial \phi(\mathbf{r})/\partial r_i))dt + (\partial D_i(\mathbf{r})/\partial r_i)dt)^2}{4D_i(\mathbf{r})dt} \right. \\ &\quad \left. + \frac{dt}{2} \frac{\partial}{\partial r_i} \left(A_i(\mathbf{r}) - 2\frac{\partial D_i(\mathbf{r})}{\partial r_i} + 2D_i(\mathbf{r})\frac{\partial \phi(\mathbf{r})}{\partial r_i} \right) + \frac{dt}{4} \frac{\partial^2 D_i(\mathbf{r})}{\partial r_i^2} \right]. \end{aligned} \quad (44)$$

Since we have in both cases chosen evaluation at $a = b = 1/2$ we note that multiplication follows Stratonovich rules so that we have $f(\mathbf{r})dx_i = f(\mathbf{x}) \circ dx_i$. Considering

the ratio of these two propagators we find

$$\begin{aligned} d\Delta S_{\text{ex}} &= \ln \frac{p(\mathbf{x}', t + dt | \mathbf{x}, t)}{p^{\text{ad}}(\mathbf{x}, t + dt | \mathbf{x}', t)} = \\ &\quad \sum_i D_i(\mathbf{x}) \left(\frac{\partial \phi(\mathbf{x})}{\partial x_i} \right)^2 dt + A_i(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_i} dt \\ &\quad - \frac{\partial \phi(\mathbf{x})}{\partial x_i} \circ dx_i - \frac{\partial A_i(\mathbf{x})}{\partial x_i} dt + \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} dt \\ &\quad - D_i(\mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_i^2} dt - 2 \frac{\partial D_i(\mathbf{x})}{\partial x_i} \frac{\partial \phi(\mathbf{x})}{\partial x_i} dt. \end{aligned} \quad (45)$$

However, we also have the condition

$$\begin{aligned} \nabla \cdot J^{\text{st}}(\mathbf{x}) &= 0 \\ &= \sum_i \frac{\partial}{\partial x_i} \left(e^{-\phi(\mathbf{x})} \left(A_i(\mathbf{x}) - \frac{\partial D_i(\mathbf{x})}{\partial x_i} + D_i(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_i} \right) \right) \\ &= \left(-A_i(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_i} - D_i(\mathbf{x}) \left(\frac{\partial \phi(\mathbf{x})}{\partial x_i} \right)^2 + \frac{\partial A_i(\mathbf{x})}{\partial x_i} \right. \\ &\quad \left. - \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} + D_i(\mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_i^2} + 2 \frac{\partial D_i(\mathbf{x})}{\partial x_i} \frac{\partial \phi(\mathbf{x})}{\partial x_i} \right) e^{-\phi(\mathbf{x})} \end{aligned} \quad (46)$$

and so by insertion we arrive at

$$\ln \frac{p(\mathbf{x}', t + dt | \mathbf{x}, t)}{p^{\text{ad}}(\mathbf{x}, t + dt | \mathbf{x}', t)} = \sum_i - \frac{\partial \phi(\mathbf{x})}{\partial x_i} \circ dx_i \quad (47)$$

which justifies the usual characterisation of the adjoint dynamics [15, 19, 24] for use in continuous dynamics when written

$$\frac{p(\mathbf{x}', t + dt | \mathbf{x}, t)}{p^{\text{ad}}(\mathbf{x}, t + dt | \mathbf{x}', t)} = \frac{p^{\text{st}}(\mathbf{x}')}{p^{\text{st}}(\mathbf{x})} \quad (48)$$

through consideration of Eq. (38) and the Stratonovich rules which mimic normal calculus.

We construct an increment in ΔS_1 , using the above result with the inclusion of a change in system entropy such that

$$\begin{aligned} d\Delta S_1 &= d\Delta S_{\text{sys}} + d(\Delta Q_{\text{ex}}/k_B T_{\text{env}}) \\ &= -d(\ln p) - \sum_i \frac{\partial \phi(\mathbf{x})}{\partial x_i} \circ dx_i \\ &= -\frac{1}{p(\mathbf{x})} \frac{\partial p(\mathbf{x})}{\partial t} dt - \sum_i \frac{1}{p(\mathbf{x})} \frac{\partial p(\mathbf{x})}{\partial x_i} dx_i \\ &\quad - D_i(\mathbf{x}) \left(\frac{1}{p(\mathbf{x})} \frac{\partial^2 p(\mathbf{x})}{\partial x_i^2} - \frac{1}{(p(\mathbf{x}))^2} \left(\frac{\partial p(\mathbf{x})}{\partial x_i} \right)^2 \right) dt \\ &\quad - \frac{\partial \phi(\mathbf{x})}{\partial x_i} dx_i - D_i(\mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_i^2} dt. \end{aligned} \quad (49)$$

Applying the same averaging procedure used to calculate $\langle d\Delta S_{\text{tot}} \rangle$ we find

$$\begin{aligned} \langle d\Delta S_1 \rangle &= \sum_i \int d\mathbf{x} p(\mathbf{x}) \frac{\partial A_i(\mathbf{x})}{\partial x_i} dt - \frac{\partial^2(D_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i^2} dt \\ &\quad + \frac{D_i(\mathbf{x})}{p(\mathbf{x})} \left(\frac{\partial p(\mathbf{x})}{\partial x_i} \right)^2 dt - D_i(\mathbf{x}) \frac{\partial^2 p(\mathbf{x})}{\partial x_i^2} dt \\ &\quad - p(\mathbf{x}) A(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_i} dt - p(\mathbf{x}) D(\mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_i^2} dt. \end{aligned} \quad (50)$$

However, using Eq. (46) we may represent this as

$$\begin{aligned} \langle d\Delta S_1 \rangle &= \\ &\sum_i \int d\mathbf{x} \frac{D_i(\mathbf{x})}{p(\mathbf{x})} \left(\frac{\partial p(\mathbf{x})}{\partial x_i^2} \right)^2 dt + p(\mathbf{x}) D_i(\mathbf{x}) \left(\frac{\partial \phi(\mathbf{x})}{\partial x_i^2} \right)^2 dt \\ &\quad - 2p(\mathbf{x}) \frac{\partial D_i(\mathbf{x})}{\partial x_i} \frac{\partial \phi(\mathbf{x})}{\partial x_i} - 2p(\mathbf{x}) D_i(\mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_i^2} dt \\ &\quad + p(\mathbf{x}) \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} dt - \frac{\partial^2(D_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i^2} dt - D_i(\mathbf{x}) \frac{\partial^2 p(\mathbf{x})}{\partial x_i^2} dt. \end{aligned} \quad (51)$$

By further integration by parts, dropping boundary terms and rearranging, this becomes

$$\begin{aligned} \langle d\Delta S_1 \rangle &= \sum_i \int d\mathbf{x} \frac{D_i(\mathbf{x})}{p(\mathbf{x})} \left(\frac{\partial p(\mathbf{x})}{\partial x_i^2} \right)^2 dt \\ &\quad + p(\mathbf{x}) D_i(\mathbf{x}) \left(\frac{\partial \phi(\mathbf{x})}{\partial x_i^2} \right)^2 dt + 2D_i(\mathbf{x}) \frac{\partial p(\mathbf{x})}{\partial x_i} \frac{\partial \phi(\mathbf{x})}{\partial x_i} dt \end{aligned} \quad (52)$$

which can be written

$$\begin{aligned} \frac{d\langle \Delta S_1 \rangle}{dt} &= \frac{d\langle \Delta S_{\text{sys}} + (\Delta Q_{\text{ex}}/k_B T_{\text{env}}) \rangle}{dt} \\ &= \sum_i \int d\mathbf{x} \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i(\mathbf{x})}{p(\mathbf{x})} - \frac{J_i^{\text{st}}(\mathbf{x})}{p^{\text{st}}(\mathbf{x})} \right)^2 \\ &= \sum_i \int d\mathbf{x} \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i^{\text{ir}}(\mathbf{x})}{p(\mathbf{x})} - \frac{J_i^{\text{st, ir}}(\mathbf{x})}{p^{\text{st}}(\mathbf{x})} \right)^2 \end{aligned} \quad (53)$$

assuring the positivity of such a contribution. Since it can be written in terms of the total current J_i in this way, it maps precisely onto the non-adiabatic entropy production appearing in [21] and thus can be expressed as

$$\frac{d\langle \Delta S_1 \rangle}{dt} = - \int d\mathbf{x} \frac{\partial p(\mathbf{x})}{\partial t} \ln \frac{p(\mathbf{x})}{p^{\text{st}}(\mathbf{x})} \quad (54)$$

as highlighted by the authors of [21]. We however emphasise that Eq. (54) is to be considered alongside the accompanying SDE in Eq. (49), from which it has been derived directly, rather than by a division of an observed positive contribution to the mean rate of change of Gibbs entropy into presumed unique transient and stationary terms.

We may now by similar means consider an increment in ΔS_2 as follows:

$$d\Delta S_2 = \ln \frac{p(\mathbf{x}', t + dt | \mathbf{x}, t)}{p^{\text{ad}}(\varepsilon \mathbf{x}', t + dt | \varepsilon \mathbf{x}, t)}. \quad (55)$$

In this case the construction of the denominator follows slightly different rules since, unlike ΔS_{tot} and ΔS_1 , the

alternative path, $\vec{\mathbf{x}}^T$, is based on a time reversal of the coordinates, but otherwise follows the sequence of the forward path. As such b behaves in the same manner as a rendering $r'_i = bx'_i + (1 - b)x_i$, $f(\mathbf{r}')dx_i = f(\mathbf{x})dx_i + 2bD(\mathbf{x})\partial f(\mathbf{x})/\partial x_i dt$, as detailed in appendix A. In this case the appropriate choice for the equivalence of evaluation points $\mathbf{r}' = \mathbf{r}$ is $a = b$. For continuity, we

may once again choose $a = b = 1/2$ with Stratonovich multiplication rules: we represent the transition probability appearing in the numerator through Eq. (23), and the denominator by a similar means using the drift term given in Eq. (42) and the path choice $\mathbf{x}^T(t) = \varepsilon \mathbf{x}(t)$, such that

$$\begin{aligned} p^{\text{ad}}(\varepsilon \mathbf{x}', t + dt | \varepsilon \mathbf{x}, t) &= \prod_i \sqrt{\frac{1}{4\pi D_i(\varepsilon \mathbf{r})dt}} \\ &\times \exp \left[-\frac{(\varepsilon_i dx_i + (A_i(\varepsilon \mathbf{r}) - 2(\partial D_i(\varepsilon \mathbf{r})/\partial(\varepsilon_i r_i)) + 2D_i(\varepsilon \mathbf{r})(\partial \phi(\varepsilon \mathbf{r})/\partial(\varepsilon_i r_i)))dt + (\partial D_i(\varepsilon \mathbf{r})/\partial(\varepsilon_i r_i))dt)^2}{4D_i(\varepsilon \mathbf{r})dt} \right. \\ &\quad \left. + \frac{dt}{2} \frac{\partial}{\partial \varepsilon_i r_i} \left(A_i(\varepsilon \mathbf{r}) - 2 \frac{\partial D_i(\varepsilon \mathbf{r})}{\partial(\varepsilon_i r_i)} + 2D_i(\varepsilon \mathbf{r}) \frac{\partial \phi(\varepsilon \mathbf{r})}{\partial(\varepsilon_i r_i)} \right) + \frac{dt}{4} \frac{\partial^2 D_i(\varepsilon \mathbf{r})}{\partial(\varepsilon_i r_i)^2} \right]. \end{aligned} \quad (56)$$

We can utilise the usual transformation rules and assumptions for A^{ir} , A^{rev} and D_i and express $\partial \phi(\varepsilon \mathbf{r})/\partial(\varepsilon_i r_i) = \varepsilon_i \partial \phi(\varepsilon \mathbf{r})/\partial r_i = \phi'_i(\varepsilon \mathbf{r})$ (along with $\partial^2 \phi(\varepsilon \mathbf{r})/\partial(\varepsilon_i r_i)^2 = \phi''_i(\varepsilon \mathbf{r})$) such that we can write the propagator as

$$\begin{aligned} p^{\text{ad}}(\varepsilon \mathbf{x}', t + dt | \varepsilon \mathbf{x}, t) &= \prod_i \sqrt{\frac{1}{4\pi D_i(\mathbf{r})dt}} \\ &\times \exp \left[-\frac{(dx_i + (A_i^{\text{ir}}(\mathbf{r}) - A_i^{\text{rev}}(\mathbf{r}) - 2(\partial D_i(\mathbf{r})/\partial r_i) + 2\varepsilon_i D_i(\mathbf{r})\phi'_i(\varepsilon \mathbf{r}))dt + (\partial D_i(\mathbf{r})/\partial r_i)dt)^2}{4D_i(\mathbf{r})dt} \right. \\ &\quad \left. + \frac{dt}{2} \frac{\partial}{\partial r_i} \left(A_i^{\text{ir}}(\mathbf{r}) - A_i^{\text{rev}}(\mathbf{r}) - 2 \frac{\partial D_i(\mathbf{r})}{\partial r_i} + 2\varepsilon_i D_i(\mathbf{r})\phi'_i(\varepsilon \mathbf{r}) \right) + \frac{dt}{4} \frac{\partial^2 D_i(\mathbf{r})}{\partial r_i^2} \right]. \end{aligned} \quad (57)$$

Constructing the ratio in Eq. (55) we find

$$\begin{aligned} d\Delta S_2 &= d(\Delta Q_{\text{hk,G}}/k_B T_{\text{env}}) \\ &= \sum_i -\frac{A_i^{\text{ir}}(\mathbf{x})A_i^{\text{rev}}(\mathbf{x})}{D_i(\mathbf{x})} dt + \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dx_i \\ &\quad + \varepsilon_i \phi'_i(\varepsilon \mathbf{x}) dx_i - \frac{1}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x} dx_i \\ &\quad + \frac{1}{D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 dt + D_i(\mathbf{x})(\phi'_i(\varepsilon \mathbf{x}))^2 dt \\ &\quad - 2\varepsilon_i \phi'_i(\varepsilon \mathbf{x}) \frac{\partial D_i(\mathbf{x})}{\partial x} dt + \varepsilon_i (A_i^{\text{ir}}(\mathbf{x}) - A_i^{\text{rev}}(\mathbf{x})) \phi'_i(\varepsilon \mathbf{x}) dt \\ &\quad - \frac{(A_i^{\text{ir}}(\mathbf{x}) - A_i^{\text{rev}}(\mathbf{x}))}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dt. \end{aligned} \quad (59)$$

which is the same as

$$\begin{aligned} d\Delta S_2 &= \sum_i -\frac{A_i^{\text{ir}}(\mathbf{x})A_i^{\text{rev}}(\mathbf{x})}{D_i(\mathbf{x})} dt + \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} \circ dx_i \\ &\quad - \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} dt + D_i(\mathbf{x})(\phi'_i(\varepsilon \mathbf{x}))^2 dt - 2\varepsilon_i \frac{\partial D_i(\mathbf{x})}{\partial x_i} \phi'_i(\varepsilon \mathbf{x}) dt \\ &\quad + \varepsilon_i (A_i^{\text{ir}}(\mathbf{x}) - A_i^{\text{rev}}(\mathbf{x})) \phi'_i(\varepsilon \mathbf{x}) dt + \frac{A_i^{\text{rev}}(\mathbf{x})}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dt \\ &\quad + \varepsilon_i \phi'_i(\varepsilon \mathbf{x}) \circ dx_i - \frac{1}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x} \circ dx_i \\ &\quad + \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} dt - D_i(\mathbf{x}) \phi''_i(\varepsilon \mathbf{x}) dt \end{aligned} \quad (58)$$

By employing the averaging procedure we find

$$\begin{aligned} \frac{d\langle \Delta S_2 \rangle}{dt} &= \\ &\sum_i \int d\mathbf{x} \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(A_i^{\text{ir}}(\mathbf{x}) - \frac{\partial D_i(\mathbf{x})}{\partial x_i} + \varepsilon_i D_i(\mathbf{x}) \phi'_i(\varepsilon \mathbf{x}) \right)^2 \end{aligned} \quad (60)$$

which may be written

$$\begin{aligned} \frac{d\langle\Delta S_2\rangle}{dt} &= \frac{d\langle\Delta Q_{hk,G}/k_B T_{env}\rangle}{dt} \\ &= \sum_i \int d\mathbf{x} \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i^{ir,st}(\varepsilon\mathbf{x})}{p^{st}(\varepsilon\mathbf{x})} \right)^2. \end{aligned} \quad (61)$$

Such a form illustrates the positivity requirement of ΔS_2 in the mean, resulting from its adherence to an IFT, and again Eq. (61) is to be considered alongside the complementary SDE in Eq. (59). Since it is based on an integral over the stationary irreversible flux, $d\langle\Delta S_2\rangle/dt$ describes a contribution to entropy production which arises from an absence of detailed balance and is non-zero both in and out of stationarity. This quantity is to be contrasted with the adiabatic entropy production in [21] which we may now consider to be a special case when there are only even variables in the dynamics. We point out again the importance of the direct derivation of this result from the SDE in this formalism, as opposed to a division of the irreversible flux into terms with structure based solely on $p^{st}(\mathbf{x})$, which would not have obviously led to the above expression.

We note that the integral in Eq. (61) must reduce to the total entropy production, and thus an integral over the stationary irreversible flux (i.e. without the ε factors inside the squared term in the integrand), in the stationary state, but there are other circumstances when this correspondence applies more generally. A first case is when the irreversible stationary flux is proportional to the stationary distribution, which would be the case for a non-equilibrium constraint that is independent of the phase space variables, as illustrated later in example II, and a second case is when the total flux in each coordinate is everywhere zero ($J_i(\mathbf{x}) = 0$), such as for independent variables, x_i , defined on regions with natural or reflecting boundaries.

To complete the description of all three contributions to entropy production we now consider an increment in ΔS_3 . By using the definition in Eq. (14)

$$d\Delta S_3 = \ln \frac{p^{\text{ad}}(\varepsilon\mathbf{x}', t + dt|\varepsilon\mathbf{x}, t)p^{\text{ad}}(\mathbf{x}, t + dt|\mathbf{x}', t)}{p(\mathbf{x}', t + dt|\mathbf{x}, t)p(\varepsilon\mathbf{x}, t + dt|\varepsilon\mathbf{x}', t)} \quad (62)$$

together with the previously used propagators, and employing the stationarity condition evaluated at $\varepsilon\mathbf{x}$:

$$\begin{aligned} \nabla \cdot J^{\text{st}}(\varepsilon\mathbf{x}) &= 0 \\ &= \sum_i \left(-(A_i^{\text{ir}}(\varepsilon\mathbf{x}) + A_i^{\text{rev}}(\varepsilon\mathbf{x}))\phi'_i(\varepsilon\mathbf{x}) + \frac{\partial A_i^{\text{ir}}(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)} \right. \\ &\quad \left. + \frac{\partial A_i^{\text{rev}}(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)} - D_i(\varepsilon\mathbf{x})(\phi'_i(\varepsilon\mathbf{x}))^2 \right. \\ &\quad \left. - \frac{\partial^2 D_i(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)^2} + D_i(\varepsilon\mathbf{x})\phi''_i(\varepsilon\mathbf{x}) + 2\frac{\partial D_i(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)}\phi'_i(\varepsilon\mathbf{x}) \right) e^{-\phi(\varepsilon\mathbf{x})} \end{aligned} \quad (63)$$

we find

$$\begin{aligned} d\Delta S_3 &= d(\Delta Q_{hk,T}/k_B T_{env}) \\ &= \sum_i \phi'_i(\mathbf{x}) \circ dx_i - \varepsilon_i \phi'_i(\varepsilon\mathbf{x}) \circ dx_i \\ &= \sum_i \ln \frac{\exp[-\phi(\mathbf{x})]}{\exp[-\phi(\mathbf{x}')]} \frac{\exp[-\phi(\varepsilon\mathbf{x}')]}{\exp[-\phi(\varepsilon\mathbf{x})]}. \end{aligned} \quad (64)$$

which maps onto the same quantity derived from a master equation approach [23]. We can then construct the average contribution by converting to Ito form and performing the path integral such that

$$\begin{aligned} \langle d\Delta S_3 \rangle &= \sum_i \int d\mathbf{x} p(\mathbf{x}) A_i(\mathbf{x}) (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon\mathbf{x})) dt \\ &\quad + p(\mathbf{x}) D_i(\mathbf{x}) (\phi''_i(\mathbf{x}) - \phi''_i(\varepsilon\mathbf{x})) dt \end{aligned} \quad (65)$$

and proceed to manipulate by integrating by parts, assuming the probability density and current vanish or cancel at boundaries, such that

$$\begin{aligned} \langle d\Delta S_3 \rangle &= \sum_i \int d\mathbf{x} p(\mathbf{x}) A_i(\mathbf{x}) (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon\mathbf{x})) dt \\ &\quad - \int d\mathbf{x} \frac{\partial}{\partial x_i} (p(\mathbf{x}) D_i(\mathbf{x})) (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon\mathbf{x})) dt \\ &= \sum_i \int d\mathbf{x} (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon\mathbf{x})) \\ &\quad \times \left(A_i(\mathbf{x}) p(\mathbf{x}) - \frac{\partial}{\partial x_i} (p(\mathbf{x}) D_i(\mathbf{x})) \right) dt \\ &= \sum_i \int d\mathbf{x} (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon\mathbf{x})) J_i(\mathbf{x}) dt \\ &\quad - \int d\mathbf{x} (\phi(\mathbf{x}) - \phi(\varepsilon\mathbf{x})) \frac{\partial J_i(\mathbf{x})}{\partial x_i} dt \\ &= - \int d\mathbf{x} (\phi(\mathbf{x}) - \phi(\varepsilon\mathbf{x})) \left(\sum_i \frac{\partial J_i(\mathbf{x})}{\partial x_i} \right) dt \\ &= - \int d\mathbf{x} (\phi(\mathbf{x}) - \phi(\varepsilon\mathbf{x})) (\nabla \cdot J(\mathbf{x})) dt. \end{aligned} \quad (66)$$

By substituting the original Fokker-Planck equation we may also write this as

$$\begin{aligned} \frac{d\langle\Delta S_3\rangle}{dt} &= \frac{d\langle\Delta Q_{hk,T}/k_B T_{env}\rangle}{dt} \\ &= \int d\mathbf{x} \frac{\partial p(\mathbf{x})}{\partial t} (\phi(\mathbf{x}) - \phi(\varepsilon\mathbf{x})) \\ &= - \int d\mathbf{x} \frac{\partial p(\mathbf{x})}{\partial t} \ln \frac{p^{\text{st}}(\mathbf{x})}{p^{\text{st}}(\varepsilon\mathbf{x})}. \end{aligned} \quad (67)$$

This has a form similar to Eq. (54) and is clearly a contribution to the mean total entropy production rate that behaves transiently in a manner similar to ΔS_1 . The quantity ΔS_1 appears in the Hatano-Sasa relation which describes the entropy production associated with a transition between different stationary states. However, in

light of Eq. (67) we suggest that ΔS_1 , and thus the Hatano-Sasa relation and non-adiabatic entropy production, do not represent the entire entropy production associated with transitions between stationary states (or more generally relaxation) since, in the mean, we can construct a new quantity which comprises all contributions which are non-zero only during relaxation, by combining Eqs. (54) and (67) giving

$$\begin{aligned} \frac{d\langle\Delta S_1 + \Delta S_3\rangle}{dt} &= \frac{d\langle\Delta S_{\text{sys}} + (\Delta Q_{\text{ex}} + \Delta Q_{\text{hk,T}})/k_B T_{\text{env}}\rangle}{dt} \\ &= - \int d\mathbf{x} \frac{\partial p(\mathbf{x})}{\partial t} \ln \frac{p(\mathbf{x})}{p^{\text{st}}(\varepsilon\mathbf{x})}. \end{aligned} \quad (68)$$

This describes a contribution to the mean entropy production rate which occurs when the system is out of stationarity, but it does not obey an IFT and thus has no guarantee of positivity.

Our central results therefore are expressions for three contributions to entropy production for arbitrary systems with odd and even dynamical variables evolving according to Ito SDEs with multiplicative noise. These expressions apply to individual trajectories (Eqs. (31), (49), (59) and (64)) and in the mean (Eqs. (37), (53), (61) and (67)). Such a demonstration shows the additional complexity introduced by the inclusion of odd variables if one insists on considering entropy production to be due to relaxation or to non-equilibrium constraints with particular reference to Eq. (68). One may think of $\langle\Delta S_1 + \Delta S_3\rangle$ as describing a transient contribution to entropy production in the same manner as $\langle\Delta S_1\rangle$, but with the further specification of the nature of the coordinates: the entropy production depends on whether the variables being described are odd or even. The additional complexity of ΔS_3 arises because Eq. (68) can only differ from Eq. (54) when the stationary state is out of equilibrium, such that $p^{\text{st}}(\mathbf{x}) \neq p^{\text{st}}(\varepsilon\mathbf{x})$.

We of course expect and require that the contributions detailed here are related such that

$$\frac{d\langle\Delta S_{\text{tot}}\rangle}{dt} = \frac{d\langle\Delta S_1\rangle}{dt} + \frac{d\langle\Delta S_2\rangle}{dt} + \frac{d\langle\Delta S_3\rangle}{dt} \quad (69)$$

yet their forms derived above do not obviously lend themselves to such a demonstration immediately. For completeness this is shown in appendix B.

V. EXAMPLE I: STATIONARY HEAT TRANSPORT

We provide as a first example of usage of the above formalism a physical situation which necessitates the use of odd variables in order to describe entropy production adequately: heat transport due to diffusion in one spatial dimension in the presence of a spatially dependent temperature field. Mathematically this system may be

modelled without odd (velocity) variables by employing the overdamped limit and constructing a multiplicative SDE and Fokker-Planck equation of the form

$$dx = \frac{F(x)}{m\gamma} dt + \sqrt{\frac{2k_B T(x)}{m\gamma}} dW \quad (70)$$

and

$$\frac{\partial p(x,t)}{\partial t} = - \frac{\partial}{\partial x} \left(\frac{F(x)p(x,t)}{m\gamma} \right) + \frac{\partial^2}{\partial x^2} \left(\frac{k_B T(x)p(x,t)}{m\gamma} \right) \quad (71)$$

where m is the particle mass, γ the damping coefficient and $F(x)$ the force operating on the particle. We note the Ito form of both (for a discussion of the resolution of the Ito-Stratonovich dilemma in this case see, for example, [29–31]). This Fokker-Planck equation has a stationary distribution

$$p^{\text{st}}(x) = \frac{\mathcal{N}m\gamma}{k_B T(x)} \exp \left[\int_0^x dx' \frac{F(x')}{k_B T(x')} \right] \quad (72)$$

where \mathcal{N} is a normalisation constant. We can quite readily identify the terms $A_x^{\text{ir}} = F(x)/m\gamma$, $A_x^{\text{rev}} = 0$ and $D_x(x) = k_B T(x)/m\gamma$. However, when we come to construct the dimensionless entropy production in the stationary state from Eq. (31) as

$$\begin{aligned} d\Delta S_{\text{tot}} &= \\ &\frac{A_x^{\text{ir}}(x)}{D_x(x)} \circ dx - \frac{1}{D_x(x)} \frac{\partial D_x(x)}{\partial x} \circ dx - \frac{1}{p^{\text{st}}(x)} \frac{\partial p^{\text{st}}(x)}{\partial x} \circ dx \\ &= \left[\frac{F(x)}{k_B T(x)} - \frac{1}{T(x)} \frac{\partial T(x)}{\partial x} \right. \\ &\quad \left. - \frac{1}{p^{\text{st}}(x)} \left(-\frac{1}{T(x)} \frac{\partial T(x)}{\partial x} p^{\text{st}}(x) + \frac{F(x)}{k_B T(x)} p^{\text{st}}(x) \right) \right] \circ dx \\ &= 0 \end{aligned} \quad (73)$$

we find that there is zero entropy production for all trajectories. This may be understood either physically by recognising that in the overdamped limit one demands that the velocity distribution relaxes instantaneously thereby preventing any heat transfer due to temperature inhomogeneities, or geometrically by recognising the impossibility of having stationary flow, and thus entropy production, for a system in one dimension with natural boundaries.

To provide a satisfactory representation and to understand the entropy production in such a system we need to consider the more realistic underdamped dynamics in full phase space where we retain both position and velocity coordinates, x and v , which are even and odd under time reversal, respectively. The SDEs and Fokker-Planck equation are now given as

$$\begin{aligned} dx &= vdt \\ dv &= -\gamma v dt + \frac{F(x)}{m} dt + \sqrt{\frac{2k_B T(x)\gamma}{m}} dW \end{aligned} \quad (74)$$

and

$$\begin{aligned} \frac{\partial p(x, v, t)}{\partial t} &= -v \frac{\partial p(x, v, t)}{\partial x} \\ &- \frac{\partial}{\partial v} \left(\left(\frac{F(x)}{m} - \gamma v \right) p(x, v, t) \right) + \frac{k_B T(x) \gamma}{m} \frac{\partial^2 p(x, v, t)}{\partial v^2}. \end{aligned} \quad (75)$$

We may then identify the terms $A_x^{\text{ir}} = 0$, $A_x^{\text{rev}} = v$, $A_v^{\text{ir}} = -\gamma v$, $A_v^{\text{rev}} = F(x)/m$, $D_x = 0$ and $D_v = k_B T(x) \gamma / m$. By Eq. (31) the dimensionless entropy production is

$$\begin{aligned} d\Delta S_{\text{tot}} &= -d(\ln p(x, v, t)) - \frac{mv}{k_B T(x)} \circ dv + \frac{Fv}{k_B T(x)} dt \\ &= -d(\ln p(x, v, t)) - \frac{1}{k_B T(x)} d \left(\frac{mv^2}{2} \right) + \frac{F}{k_B T(x)} dx \end{aligned} \quad (76)$$

using $v \circ dv = (1/2)(v' + v)(v' - v)$ and $vdt = dx$, and noting that x is now deterministic, meaning the integration rules are irrelevant. The second and third terms correctly reproduce the form of the change in medium entropy as heat transfer to the environment, equal to negative heat transfer to the particle (in agreement with the result found in stochastic energetics [32]), divided by the instantaneous temperature, and do so only by virtue of the consideration of odd and even variables.

We can use this SDE to produce distributions of entropy production and verify relevant fluctuation theorems. To do so, however, requires knowledge of the solution to the Fokker-Planck equation, for which there is no simple analytical form. To proceed we restrict ourselves to the stationary state and utilise the expansion found in [31] and [33] which expresses the stationary solution as a series expansion about the overdamped distribution:

$$\begin{aligned} p^{\text{st,over}}(x, v) &= \frac{\mathcal{N}m}{k_B T(x)} \exp \left[\int_0^x dx' \frac{F(x')}{k_B T(x')} \right] \\ &\times \sqrt{\frac{m}{2\pi k_B T(x)}} \exp \left[-\frac{mv^2}{2k_B T(x)} \right], \end{aligned} \quad (77)$$

where \mathcal{N} is determined by normalisation, such that

$$p^{\text{st}}(x, v) = p^{\text{st,over}}(x, v) + \sum_{i=1}^{\infty} (1/\gamma)^i p_i(x, v) \quad (78)$$

$p_i(x, v)$ has a general form

$$\begin{aligned} p_i(x, v) &= \\ &\sum_{k=a_i}^{k=b_i} \frac{c_{i,k}(x) H_k(v \sqrt{m/k_B T(x)})}{\sqrt{2\pi k_B T(x)/m}} \exp \left[-\frac{mv^2}{2k_B T(x)} \right], \end{aligned} \quad (79)$$

where constants a_i , b_i and functions $c_{i,k}(x)$ are found by an iterative procedure, and $H_k(y)$ are Hermite polynomials defined as

$$H_k(y) = (-1)^k e^{\frac{y^2}{2}} \frac{d^k}{dy^k} e^{-\frac{y^2}{2}}. \quad (80)$$

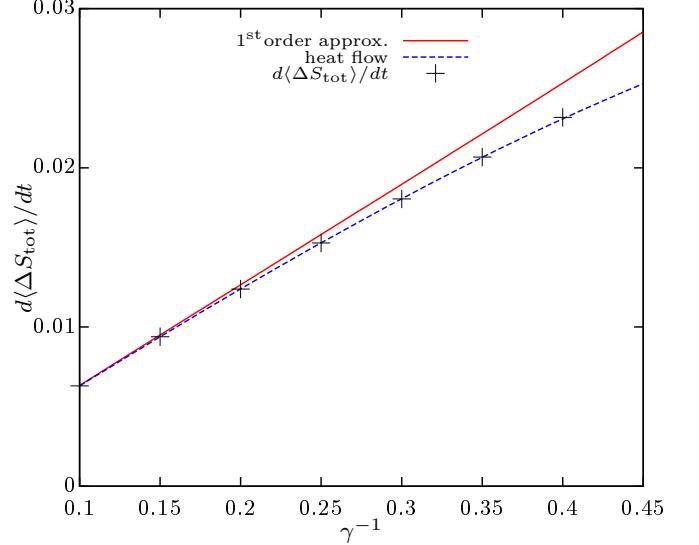


FIG. 1. Mean dimensionless entropy production, for example I, for a range of damping coefficients as predicted by a first order approximation in Eq. (84) (solid line), an integral over the heat current, Eq. (82) (dashed line) and a Monte Carlo average based on the SDE in Eq. (76) (crosses).

Whilst the expansion has the formal deficiency that the expansion parameter is not unitless it suffices for a theoretical illustration where we can consider it in a limit where it is appropriate. We consider units $k_B = T = m = 1$, a harmonic confining potential such that $F(x) = -x$, a temperature profile

$$T(x) = 1 + \frac{1}{2} \tanh(x), \quad (81)$$

and approximate the stationary distribution by considering the expansion in Eq. (78) to fourth order in γ^{-1} , applying the formalism numerically.

We first demonstrate that this approach yields a result which maps onto the expected phenomenological expression for dimensionless internal entropy generation [34]

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = \int_{-\infty}^{+\infty} dx J_Q(x) \frac{\partial}{\partial x} \left(\frac{1}{k_B T(x)} \right) \quad (82)$$

where $J_Q(x)$ is the stationary heat current defined as

$$J_Q(x) = \int_{-\infty}^{+\infty} dv \frac{1}{2} mv^3 p^{\text{st}}(x, v). \quad (83)$$

Figure 1 shows the dimensionless entropy production obtained by performing the integral in Eq. (82) using a numerically calculated $p^{\text{st}}(x, v)$, compared with that obtained by averaging the SDE in Eq. (76) by Monte Carlo simulation of the underlying particle dynamics, for a range of damping coefficients, alongside a demonstrably positive first order approximation based on the first

correction term in Eq. (78) given by

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} \simeq \int_{-\infty}^{+\infty} dx \frac{k_B p^{\text{st}}(x)}{2m\gamma T(x)} \left(\frac{\partial T(x)}{\partial x} \right)^2 \quad (84)$$

where

$$p^{\text{st}}(x) = \int_{-\infty}^{+\infty} dv p^{\text{st}}(x, v). \quad (85)$$

Our formalism for the rate of change of $\langle \Delta S_{\text{tot}} \rangle$ agrees with Eq. (82) and both are consistent with Eq. (84) in the $\gamma \rightarrow \infty$ limit. We point out that total entropy production decreases as coupling to the environment increases which may seem counter-intuitive, but we emphasise that with increased coupling, despite greater heat transfer to and from the environment, there is highly diminished spatial heat *transport* (the latter being the cause of entropy production) as the system is brought closer to a local equilibrium.

We can use the SDE for entropy production (Eq. (76)) to move beyond a classical description of mean entropy production to one described by Jarzynski, Seifert, Sekimoto and others [8, 25, 35] where we can identify entropy generating and destroying trajectories. We can explicitly calculate the distribution of total entropy production which is shown for $\gamma = 10$ in Figure 2 for various process intervals, along with a demonstration that it adheres to an IFT throughout. Additionally, since we consider the stationary state we can demonstrate a detailed fluctuation theorem of the form $p(\Delta S_{\text{tot}})/p(-\Delta S_{\text{tot}}) = \exp(\Delta S_{\text{tot}})$ [11] as shown in Figure 3.

Finally we point out that, being in the stationary state, $d\langle \Delta S_3 \rangle/dt = 0$, but since it is a non-equilibrium stationary state that is asymmetric in the odd velocity variable we have $\Delta S_3 \neq 0$ in detail, as is clear in Eq. (64). We can demonstrate the increasing range of values of ΔS_3 as γ is reduced and the system is taken further away from local equilibrium, with its symmetric velocity distribution, by generating the distribution of ΔS_3 using Eq. (64) for a given time interval, as shown in Figure 4. Such a result highlights the fact that although a non-zero $d\langle \Delta S_3 \rangle/dt$ is only possible during relaxation as shown by Eq. (67), the specific evolution of ΔS_3 for each trajectory is brought about by non-equilibrium constraints that cause the stationary solution to depart from equilibrium.

VI. EXAMPLE II: PARTICLE DRIVEN BY A NON-CONSERVATIVE FORCE

Once again utilising the full phase space Langevin description of the dynamics, we consider diffusion of a particle on a ring driven by a spatially independent non-conservative force and spatially independent (additive)

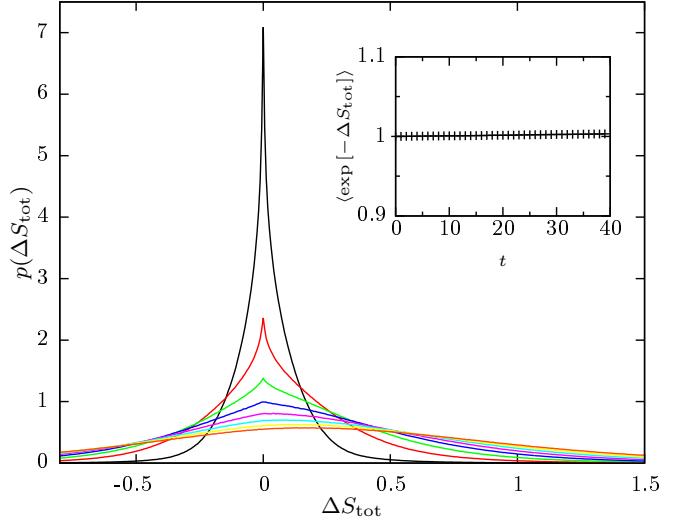


FIG. 2. Distributions of dimensionless total entropy production ΔS_{tot} , for example I, for $\gamma = 10$ together with a demonstration of adherence to an IFT. Distributions shown are for process intervals from $t = 2$ (narrowest) to $t = 44$ (widest) in steps of 6 units.

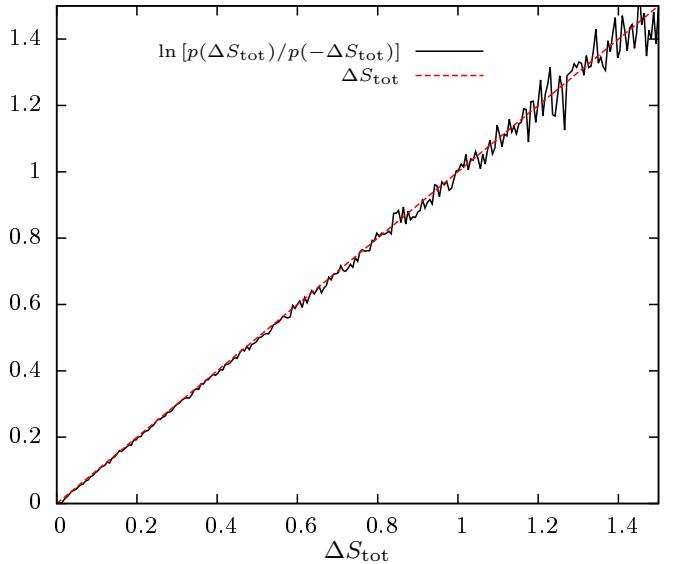


FIG. 3. Verification of a detailed fluctuation theorem for example I using data from simulation for $\gamma = 10$ at time $t = 8$.

noise such that

$$dx = vdt$$

$$dv = -\gamma v dt + \frac{F(t)}{m} dt + \sqrt{\frac{2k_B T \gamma}{m}} dW \quad (86)$$

thus giving $A_x^{\text{ir}} = 0$, $A_x^{\text{rev}} = v$, $A_v^{\text{ir}} = -\gamma v$, $A_v^{\text{rev}} = F(t)/m$, $D_x = 0$ and $D_v = k_B T \gamma / m$. For any non-zero value of $F(t)$ there will exist a stationary solution with an asymmetric Gaussian distribution in v and a uniform distribution in x due to the symmetry of the problem. Any relaxation from a given stationary state caused by

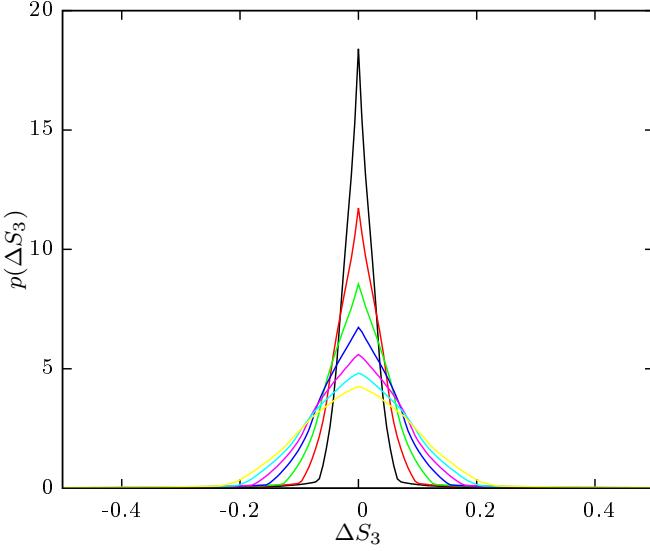


FIG. 4. Distributions of ΔS_3 for example I evaluated at $t = 8$ for a range of γ from $\gamma^{-1} = 0.1$ (narrowest) to $\gamma^{-1} = 0.4$ (widest).

changes to the non-conservative force will then also result in a uniform distribution in x for all time by the translational symmetry. As such we may proceed by considering the marginalised velocity distribution when starting from a stationary state. Exploiting the fact that the initial Gaussian solution will remain Gaussian for any $F(t)$, we can parameterise a transient solution to the Fokker-Planck equation

$$p(x, v, t) \propto \sqrt{\frac{m}{2\pi k_B T}} \exp\left[-\frac{m(v - \langle v \rangle)^2}{2k_B T}\right] \quad (87)$$

with

$$\frac{d\langle v \rangle}{dt} = \left(\frac{F}{m} - \gamma \langle v \rangle \right) \quad (88)$$

such that

$$\langle v \rangle^{\text{st}} = \frac{F}{m\gamma}. \quad (89)$$

A scenario where closed form solutions exist for all contributions to entropy production is that of an instantaneous step change in the driving force $F(t)$ so that we have

$$F(t) = \begin{cases} F_0 & t < t_0, \\ F_1 & t \geq t_0, \end{cases} \quad (90)$$

and

$$\langle v \rangle(t) = \begin{cases} F_0/m\gamma & t < t_0, \\ (F_1 + e^{-\gamma(t-t_0)}(F_0 - F_1))/m\gamma & t \geq t_0. \end{cases} \quad (91)$$

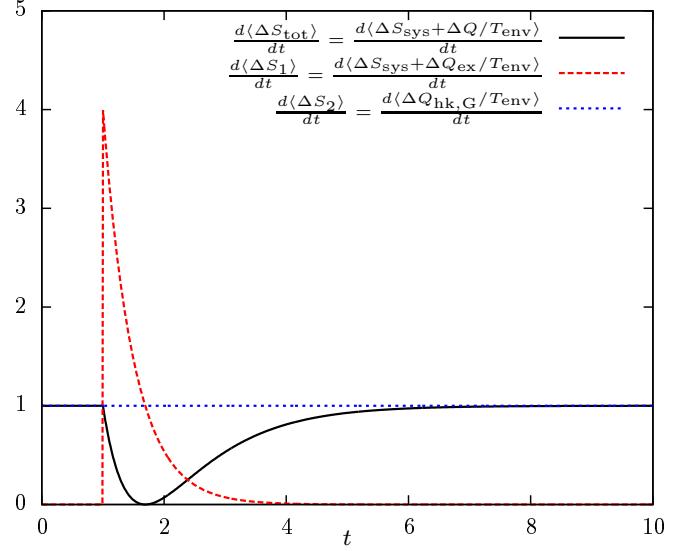


FIG. 5. Positive mean rates of dimensionless entropy change against time for example II, where we consider the transition between stationary states of a driven particle on a ring with $F_0 = 1$, $F_1 = -1$, $t_0 = 1$ and $k_B = m = \gamma = T = 1$.

Performing the relevant integrals in Eqs. (37), (53), (61) and (67) we then obtain

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = \begin{cases} F_0^2/m\gamma k_B T & t < t_0, \\ (F_0 + F_1(e^{\gamma(t-t_0)} - 1))^2 & t \geq t_0, \\ \times e^{-2\gamma(t-t_0)}/m\gamma k_B T \end{cases} \quad (92)$$

$$\frac{d\langle \Delta S_1 \rangle}{dt} = \begin{cases} 0 & t < t_0, \\ e^{-2\gamma(t-t_0)}(F_0 - F_1)^2/m\gamma k_B T & t \geq t_0, \end{cases} \quad (93)$$

$$\frac{d\langle \Delta S_2 \rangle}{dt} = \begin{cases} F_0^2/m\gamma k_B T & t < t_0, \\ F_1^2/m\gamma k_B T & t \geq t_0, \end{cases} \quad (94)$$

and

$$\frac{d\langle \Delta S_3 \rangle}{dt} = \begin{cases} 0 & t < t_0, \\ -2e^{-\gamma(t-t_0)}F_1(F_1 - F_0)/m\gamma k_B T & t \geq t_0. \end{cases} \quad (95)$$

Choosing the specific case of a reversal of the driving force such that it changes from $F_0 = 1$ to $F_1 = -1$ at time $t_0 = 1$ and employing units $k_B = m = \gamma = T = 1$, we can generate the results shown in Figures 5 and 6.

We note first that the mean rates of change of all three contributions ΔS_{tot} , ΔS_1 and ΔS_2 are positive, reflecting their adherence to an IFT. All three mean rates of change are constant for $t < t_0 = 1$, are perturbed by the change in direction of the force, and relax back to constant values consistent with the transition between the stationary states. A key feature of this behaviour is that upon perturbation, the total entropy production rate *decreases* which would not emerge

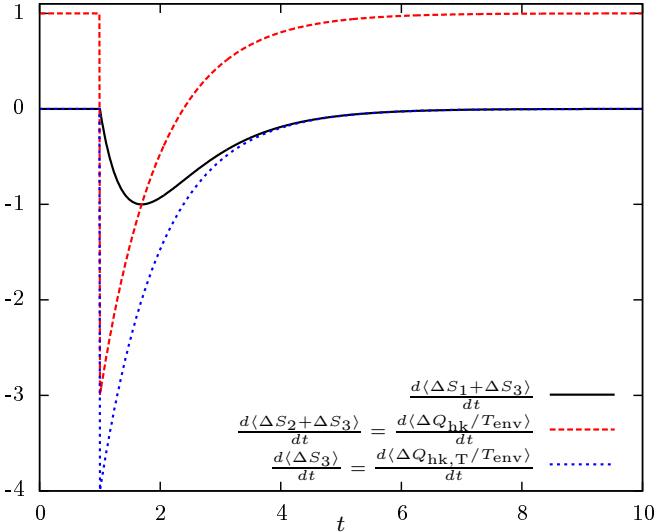


FIG. 6. Unbounded mean rates of dimensionless entropy change for example II, the driven system on a ring with $F_0 = 1$, $F_1 = -1$, $t_0 = 1$ and $k_B = m = \gamma = T = 1$.

using an overdamped description of the dynamics. This feature can be explained by the existence of the $d\langle\Delta S_3\rangle/dt$ contribution to the mean entropy production rate, which may take negative values depending on the relationship between the instantaneous distribution and the stationary distribution. In this specific case, the large negative value for $d\langle\Delta S_3\rangle/dt$ indicates that upon reversal of the force the instantaneous distribution corresponds to particle motion, on average, in a direction counter to that expected to result from the new value of the force. The velocity distribution does relax, of course, to the distribution that corresponds to the new value of the force and so the mean rate of change of ΔS_3 decays away. An important point to draw from Figure 6 is that ΔS_3 , $\Delta S_1 + \Delta S_3$ and $\Delta S_2 + \Delta S_3$ cannot be expected, in general, to be positive, reflecting that they cannot be expressed in the form of Eq. (9) and thus do not obey IFTs. This means previous approaches where the entropy production can always be divided into two positive quantities [19–21] and the house-keeping heat can be expected to obey an IFT [14], do not extend to the systems considered here.

We consider this example to be a helpful illustration of how entropy production cannot always be divided into two contributions which derive from relaxation, and an absence of detailed balance owing to a non-equilibrium constraint, respectively. Explicitly, the non-equilibrium constraint here is the constant force which produces entropy in the stationary state by inducing a constant flux around the ring. The mean rate of entropy production in that stationary state is characterised by $d\langle\Delta S_2\rangle/dt$ which remains constant throughout the process owing to the constant magnitude of the force which is applied. However, both ΔS_2 and ΔS_3 are non-zero only in the

presence of a non-equilibrium constraint which breaks detailed balance. At the same time the *mean* rate of change of ΔS_3 is non-zero only when the distribution is relaxing to a new stationary solution in the same manner as ΔS_1 . Whilst ΔS_1 describes the entropy production that arises from an evolution of the probability distribution of a general set of variables, ΔS_3 expresses what ΔS_1 explicitly leaves out: the additional impact of relaxation on entropy production that relates to the a priori physical specification of the variables as odd or even. Clearly, given that the non-equilibrium constraint is a force of constant magnitude, reflected by the constant $d\langle\Delta S_2\rangle/dt$, it is reasonable to consider the sum of ΔS_1 and ΔS_3 as the contribution that arises due to relaxation to a new stationary state, particularly when the form of its mean rate of change in Figure 6 is contrasted with that of ΔS_{tot} , ΔS_1 and ΔS_2 in Figure 5. We may make the analysis complete by considering the SDEs for all contributions. The explicit Ito forms of Eqs. (31), (49), (59) and (64) are given as

$$d\Delta S_{\text{tot}} = -\frac{m}{k_B T} \langle v \rangle dv - \frac{m}{k_B T} (v - \langle v \rangle) \frac{d\langle v \rangle}{dt} dt + \frac{F(t)}{k_B T} dx \quad (96)$$

$$d\Delta S_1 = \frac{1}{k_B T} \left(\frac{F(t)}{\gamma} - m\langle v \rangle \right) dv - \frac{m}{k_B T} (v - \langle v \rangle) \frac{d\langle v \rangle}{dt} dt \quad (97)$$

$$d\Delta S_2 = \frac{F(t)}{\gamma k_B T} dv + \frac{F(t)}{k_B T} dx \quad (98)$$

$$d\Delta S_3 = -\frac{2F(t)}{\gamma k_B T} dv \quad (99)$$

and illustrate the behaviour of all the contributions. $d\Delta S_{\text{tot}}$ is only zero when $\langle v \rangle = 0$, $F = 0$ and $d\langle v \rangle/dt = 0$ meaning the system is in the equilibrium state. $d\Delta S_1$ is zero whenever $\langle v \rangle = F/m\gamma$ and $d\langle v \rangle/dt = 0$ corresponding to any stationary state, equilibrium or otherwise, whilst $d\Delta S_2$ and $d\Delta S_3$ contribute independently of properties of the distribution (namely $\langle v \rangle$), but only when the non-equilibrium constraint is present such that $F(t) \neq 0$. $d\Delta S_3$ however, has a mean contribution of zero at stationarity since $\langle dv \rangle = 0$ for any stationary state. We can calculate distributions of all the contributions, as measured from the force reversal, numerically using the above SDEs and demonstrate the validity of IFTs, where appropriate, in Figures 7 and 8. We observe that all distributions take Gaussian form, to be expected as the model is essentially a recasting of the overdamped dragged oscillator found in [36] where the further, but non-general, detailed fluctuation theorem symmetry $p(\Delta S_{\text{tot}})/p(-\Delta S_{\text{tot}}) = \exp(\Delta S_{\text{tot}})$ has been noted to hold over finite times [36], but stressed elsewhere [37] to be coincidental. Further insight into this coincidence can be derived from the form of the SDEs which yield Gaussian distributions (for the given initial

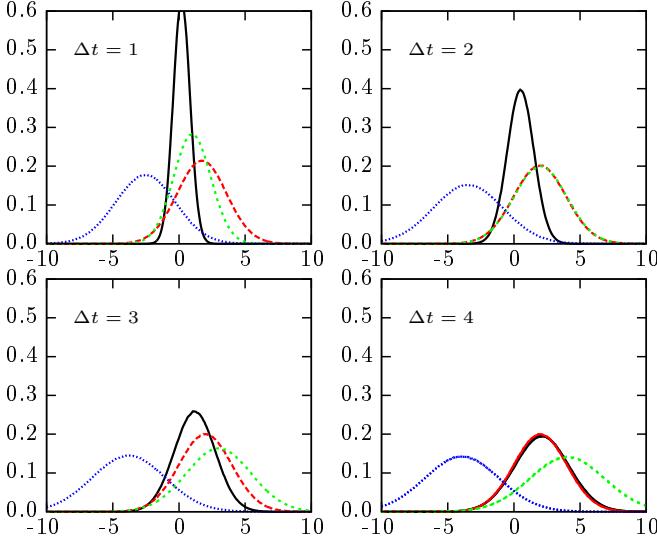


FIG. 7. Distributions of entropy productions ΔS_{tot} (solid line), ΔS_1 (wide dashed line), ΔS_2 (narrow dashed line) and ΔS_3 (dotted line) measured at times $\Delta t = t - t_0 = 1, \Delta t = 2, \Delta t = 3$ and $\Delta t = 4$ after the reversal of the force for $F_0 = 1, F_1 = -1, t_0 = 1$ and $k_B = m = \gamma = T = 1$.

conditions) since they comprise only drift and additive noise terms (that is, no terms of the form $f(v)dv$). Such properties however, do not distract from the nature of the contributions which can be readily observed: the distributions in ΔS_1 and ΔS_3 develop fastest at first reflecting the initially fast response of the distribution to the change in force. However, distributions for both ΔS_2 and ΔS_{tot} develop steadily, owing to their contributions being characterised by steady heat dissipation. As such, as time progresses, the distribution of ΔS_1 ceases to develop as the system reaches the new stationary state and the distributions of ΔS_2 and ΔS_{tot} continue to shift to the right until they eventually dominate. Similarly for ΔS_3 , we observe here that the distribution stops evolving despite receiving non-zero contributions.

For completeness we investigate the same model with a less trivial time dependence in the non-conservative force, along with its approach to the overdamped limit where such systems have been considered previously [21, 38]. We employ the force protocol

$$F(t) = 1.5 - 0.5 \tanh(-5(t - 1)) \quad (100)$$

and perform the calculations numerically for two values of damping coefficient, $\gamma = 1$ and $\gamma = 5$. We point out again that the meaning of $d\langle\Delta S_2\rangle/dt$ for this system is easily elucidated since the non-equilibrium constraint, $F(t)$, being phase space independent, leads to $J_v^{\text{ir,st}} \propto p^{\text{st}}$ so that

$$\frac{d\langle\Delta S_2\rangle}{dt} = \frac{d\langle\Delta S_{\text{tot}}\rangle^{\text{st}}}{dt} = \frac{F(t)^2}{m\gamma k_B T}. \quad (101)$$

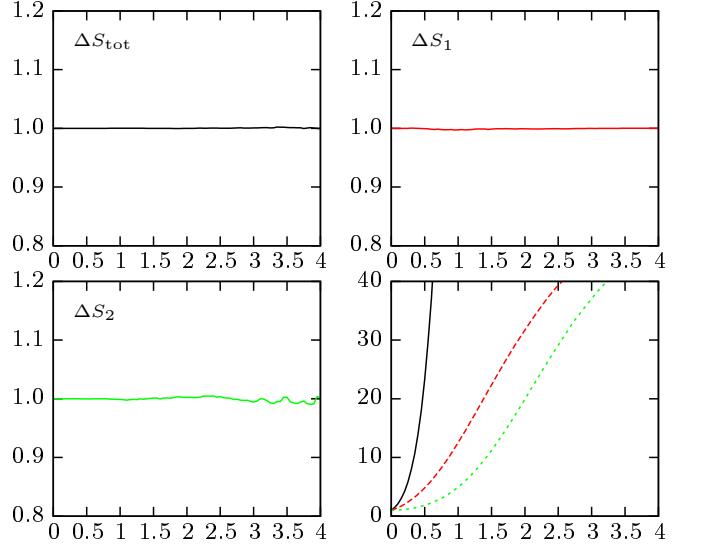


FIG. 8. Illustration of adherence to IFTs by consideration of the average $\langle \exp[-\Delta S] \rangle$ against time, $\Delta t = t - t_0$ after the force reversal, for ΔS_{tot} , ΔS_1 and ΔS_2 (indicated) and the failure to adhere to an IFT of ΔS_3 (solid line, fourth subplot), $\Delta S_1 + \Delta S_3$ (wide dashed line, fourth subplot) and $\Delta S_2 + \Delta S_3$ (narrow dashed line, fourth subplot) for $F_0 = 1, F_1 = -1, t_0 = 1$ and $k_B = m = \gamma = T = 1$.

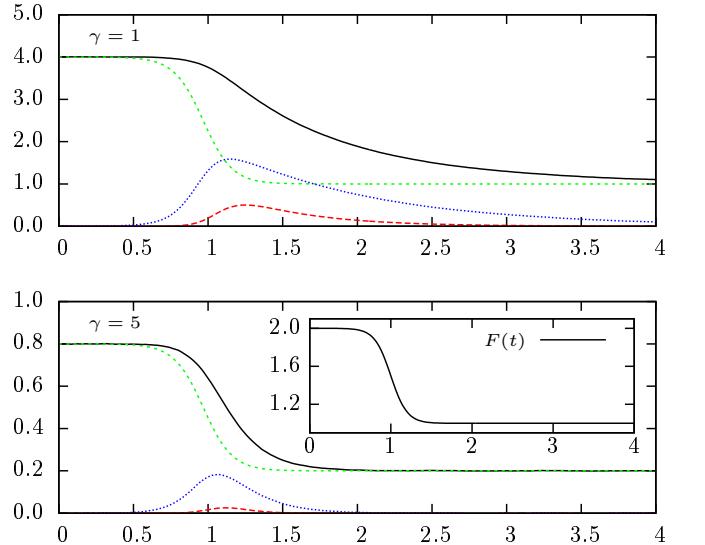


FIG. 9. Mean rates of change of ΔS_1 (wide dashed line), ΔS_2 (narrow dashed line), ΔS_3 (dotted line), and their sum ΔS_{tot} (solid line) for example II with a time dependent force given by Eq. (100) with units $k_B = m = T = 1$, and for $\gamma = 1$ (top) and $\gamma = 5$ (bottom).

The mean contributions for such a protocol for two values of the damping coefficient, again starting from the stationary state, are shown in Figure 9. Note that in this case the contribution $d\langle\Delta S_3\rangle/dt$ is positive, reflecting that as the non-conservative force decreases, the instantaneous distribution corresponds to a greater average

particle flux than would be expected from the instantaneous value of the force, thus producing more entropy in the process of relaxation from one stationary state to the other than would be expected if the relaxation were instantaneous. As γ increases, the asymmetry of the stationary state (in velocity) decreases and the contribution from ΔS_3 diminishes. Consequently, the two stationary distributions become increasingly similar, meaning the contribution ΔS_1 also diminishes rendering the total entropy production almost entirely comprised of the contribution from ΔS_2 . When the full overdamped limit is taken ΔS_2 is the only contribution and the results map onto those found in [21].

VII. DISCUSSION AND CONCLUSIONS

We have derived SDEs describing the fluctuating evolution of three contributions to entropy production, along with expressions for their mean behaviour, and demonstrated that two of these contributions obey IFTs and thus are rigorously positive in the mean. Furthermore we have demonstrated that whilst these two naturally align themselves with the irreversibility associated with relaxation, and non-equilibrium constraints, respectively, the inclusion of odd dynamical variables can give rise to a third term, which has no bounds on its sign and which cannot be so readily associated with one origin of entropy production or the other. We have sought to make these expressions as general as possible, within reason, with the intention that they may be applied to any system (physical or otherwise) described by stochastic differential equations, providing a framework for the discussion of entropy production, as defined here, within as wide a range of applications as may be relevant. To this end, we have considered a simple heat conduction problem, and after demonstrating that a full phase space representation of the dynamics is crucial to the treatment of its entropy production, we have examined specifically how it may be evaluated. The second example, that of a transition between stationary states of drift-diffusion on a ring, demonstrates the need for a third contribution to entropy production in the analysis, and provides some intuitive understanding of its nature.

We suggest that the division of the total entropy production into ΔS_1 , ΔS_2 and ΔS_3 , as we propose, is always helpful for three main reasons. The first is the ability to identify the physical origins of irreversibility in a process (relaxation and non-equilibrium constraints) and the interplay between them. The second is related to the identification of IFTs with the subsequent positivity requirements and restrictions on the statistics for the two contributions, ΔS_1 and ΔS_2 , which unambiguously align themselves with the two causes of irreversibility. And thirdly, the more delicate reason that all three contributions are constructed from total path probability

densities with equivalent measures such that they align themselves unambiguously with the same contributions found in master equation approaches which are necessarily formed from path *probabilities* [23]. This third point may be contrasted with an alternative division of the total entropy production into a system and medium contribution, neither of which can be expressed as ratios of total path probabilities. Being formed from a probability density, the system entropy, as defined in [11] and employed here (Eq. (15)) implies $S_{\text{sys}} = -\ln p(\mathbf{x}(t), t)$, is strictly not dimensionally correct (even if one argues that the relative entropy change is well defined [25]), but moreover does not share the same form as the system entropy that appears in master equation approaches, which if followed would imply $S_{\text{sys}} = -\ln p(\mathbf{x}(t), t)d\mathbf{x}(t)$. Such ambiguity then also enters into the definition of the medium entropy, but can be avoided altogether by considering the total entropy production in terms of explicit measures of irreversibility such that ΔS_{tot} is comprised of ΔS_1 , ΔS_2 and ΔS_3 , all four of which do not suffer from such issues.

Generalisations that are immediately obvious beyond the description given here may include correlated stochastic processes for which the method described should prove suitable, and processes where we no longer assume $D(\varepsilon \mathbf{x}) = D(\mathbf{x})$ for which the generalisation may be more challenging. One might note, however that processes which do not possess such a symmetry are unlikely to be physically meaningful. Furthermore, it would be natural to explore examples involving odd dynamical variables such as angular momentum and magnetic dipole moments, and to include driving by external forces that are themselves odd under time reversal, such as torques and magnetic fields. We expect to find further richness in the phenomenology of entropy production associated with stochastic dynamical behaviour.

ACKNOWLEDGMENTS

RES acknowledges financial support from the UK Engineering and Physical Sciences Research Council.

Appendix A: The use of short time propagators with multiplicative noise

Here we consider one of the terms in Eq. (24) and derive Eq. (28). By utilising Eqs. (23) and (27) we can describe an increment in the medium entropy production

according to the formalism of Seifert [11] as

$$\begin{aligned}
d\Delta S_{\text{med}} = & \sum_i \frac{1}{2} \ln D_i(\mathbf{r}') - \frac{1}{2} \ln D_i(\mathbf{r}) + \frac{dx_i^2}{4D_i(\mathbf{r}')dt} - \frac{dx_i^2}{4D_i(\mathbf{r})dt} \\
& + \frac{dx_i}{2} \left(\frac{A_i^{\text{rev}}(\mathbf{r})}{D_i(\mathbf{r})} + \frac{A_i^{\text{ir}}(\mathbf{r})}{D_i(\mathbf{r})} + \frac{A_i^{\text{ir}}(\mathbf{r}')}{D_i(\mathbf{r}')} - \frac{A_i^{\text{rev}}(\mathbf{r}')}{D_i(\mathbf{r}')} \right. \\
& \quad \left. - 2a \frac{1}{D_i(\mathbf{r})} \frac{\partial D_i(\mathbf{r})}{\partial r_i} - 2b \frac{1}{D_i(\mathbf{r}')} \frac{\partial D_i(\mathbf{r}')}{\partial r'_i} \right) \\
& - \frac{dt}{4} \left(\frac{(A_i^{\text{rev}}(\mathbf{r}) + A_i^{\text{ir}}(\mathbf{r}))^2}{D_i(\mathbf{r})} - \frac{(A_i^{\text{rev}}(\mathbf{r}') - A_i^{\text{ir}}(\mathbf{r}'))^2}{D_i(\mathbf{r}')} \right) \\
& - adt \left(D_i(\mathbf{r}) \frac{\partial}{\partial r_i} \left(\frac{A_i^{\text{ir}}(\mathbf{r})}{D_i(\mathbf{r})} \right) + D_i(\mathbf{r}) \frac{\partial}{\partial r_i} \left(\frac{A_i^{\text{rev}}(\mathbf{r})}{D_i(\mathbf{r})} \right) \right) \\
& - bdt \left(-D_i(\mathbf{r}') \frac{\partial}{\partial r'_i} \left(\frac{A_i^{\text{ir}}(\mathbf{r}')}{D_i(\mathbf{r}')}\right) + D_i(\mathbf{r}') \frac{\partial}{\partial r'_i} \left(\frac{A_i^{\text{rev}}(\mathbf{r}')}{D_i(\mathbf{r}')}\right) \right) \\
& + a^2 dt \left(\frac{\partial^2 D_i(\mathbf{r})}{\partial r_i^2} - \frac{1}{D_i(\mathbf{r})} \left(\frac{\partial D_i(\mathbf{r})}{\partial r_i^2} \right)^2 \right) \\
& - b^2 dt \left(\frac{\partial^2 D_i(\mathbf{r}')}{\partial r'^2_i} - \frac{1}{D_i(\mathbf{r}')} \left(\frac{\partial D_i(\mathbf{r}')}{\partial r'^2_i} \right)^2 \right)
\end{aligned} \tag{A1}$$

where time dependence in variables A^{ir} , A^{rev} and D_i is assumed, but not explicitly written for brevity. We may proceed by understanding that the quantity dx_i is an increment in an underlying SDE, meaning that we must consider all multiplications of the form $f(r)dx$ as infinitesimal stochastic integrals with a summation rule defined by the evaluation point r . For example, $r = x$ would imply an Ito integration, $r = (1/2)(x + x')$ would imply Stratonovich and so on. To consolidate the above it is sensible to convert all multiplications into one type, for which we choose Ito in order to apply the Ito stochastic calculus transparently using the heuristic rules $(dW_i)^2 = dt$ and $dW_i dW_j = 0$, and to drop all terms of order $dt^{3/2}$ and higher. To do so we apply the following reasoning. For a suitably smooth function, $f(\mathbf{r})$, and for infinitesimal dt , with \mathbf{r} constructed from \mathbf{x} and \mathbf{x}' using a parameter a , and for the case of a diagonal diffusion

matrix, we may write

$$\begin{aligned}
f(\mathbf{r}) &= f((1-a)\mathbf{x} + a\mathbf{x}') \\
&\simeq (1-a)f(\mathbf{x}) + af(\mathbf{x}') \\
&= f(\mathbf{x}) + a(f(\mathbf{x}') - f(\mathbf{x})) \\
&= f(\mathbf{x}) + adf(\mathbf{x}) \\
&= f(\mathbf{x}) \\
&+ a \left(\frac{\partial f(\mathbf{x})}{\partial t} dt + \nabla f(\mathbf{x}) \cdot d\mathbf{x} + \sum_i \frac{B_i(\mathbf{x})^2}{2} \nabla^2 f(\mathbf{x}) dt \right) \\
&= f(\mathbf{x}) + a \left(\frac{\partial f(\mathbf{x})}{\partial t} dt + \sum_i \frac{1}{2} B_i(\mathbf{x})^2 \nabla^2 f(\mathbf{x}) dt \right. \\
&\quad \left. + \sum_i \frac{\partial f(\mathbf{x})}{\partial x_i} (A_i(\mathbf{x}) dt + B_i(\mathbf{x}) dW_i) \right).
\end{aligned} \tag{A2}$$

Considering all instances of multiplication along with the definitions of \mathbf{r} , \mathbf{r}' , a and b , we find the following heuristic rules

$$f(\mathbf{r})dt = f(\mathbf{x})dt + O(dt^{3/2}) \tag{A3}$$

$$f(\mathbf{r}')dt = f(\mathbf{x})dt + O(dt^{3/2}) \tag{A4}$$

$$f(\mathbf{r})dx_i = f(\mathbf{x})dx_i + 2aD_i(\mathbf{x}) \frac{\partial f(\mathbf{x})}{\partial x_i} dt + O(dt^{3/2}) \tag{A5}$$

$$f(\mathbf{r}')dx_i = f(\mathbf{x})dx_i + 2(1-b)D_i(\mathbf{x}) \frac{\partial f(\mathbf{x})}{\partial x_i} dt + O(dt^{3/2}) \tag{A6}$$

giving us a method for converting all multiplications into Ito form. We use a similar reasoning to approximate

$$D_i(\mathbf{r}) \simeq D_i(\mathbf{x}) + ad(D_i(\mathbf{x})) \tag{A7}$$

$$D_i(\mathbf{r}') \simeq D_i(\mathbf{x}) + (1-b)d(D_i(\mathbf{x})) \tag{A8}$$

which along with the approximations to second order in dx_i , and therefore $d(D_i(\mathbf{x}))$, of the following form:

$$(1 + d(D_i(\mathbf{x})))^{-1} \simeq 1 - d(D_i(\mathbf{x})) + d(D_i(\mathbf{x}))^2 \tag{A9}$$

$$\ln(1 + d(D_i(\mathbf{x}))) \simeq d(D_i(\mathbf{x})) - \frac{d(D_i(\mathbf{x}))^2}{2} \tag{A10}$$

and an Ito definition of $d(D(\mathbf{x}))$, allow us to write the first four terms in Eq. (A1) to first order in dt as

$$\begin{aligned}
& \frac{1}{2} \ln D_i(\mathbf{r}') - \frac{1}{2} \ln D_i(\mathbf{r}) + \frac{dx_i^2}{4D_i(\mathbf{r}')dt} - \frac{dx_i^2}{4D_i(\mathbf{r})dt} \\
& \simeq \frac{((1-b)^2 - a^2)}{2D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 dt.
\end{aligned} \tag{A11}$$

Using the above, and the heuristic rules in Eqs. (A3-A6), we obtain

$$\begin{aligned} d\Delta S_{\text{med}} = & \sum_i \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dx_i - \frac{A_i^{\text{rev}}(\mathbf{x}) A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} dt \\ & + \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} dt - \frac{\partial A_i^{\text{rev}}(\mathbf{x})}{\partial x_i} dt \\ & + \frac{A_i^{\text{rev}}(\mathbf{x})}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dt - \frac{A_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dt \\ & - (a+b) \frac{1}{D_i(\mathbf{x})} \frac{\partial D_i(\mathbf{x})}{\partial x_i} dx_i \\ & + (b^2 - 2b - a^2) \left(\frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} - \frac{1}{D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 \right) dt \\ & + \frac{((1-b)^2 - a^2)}{2D_i(\mathbf{x})} \left(\frac{\partial D_i(\mathbf{x})}{\partial x_i} \right)^2 dt \end{aligned} \quad (\text{A12})$$

which depends on the choice of a and b . We note, however that without multiplicative noise (ie. for $(\partial/\partial x_i)D_i(\mathbf{x}) = 0$), where the inherent mathematical ambiguity in stochastic integrals is absent, the dependence on evaluation points (a and b) disappears. So this dependence is evidently related to the ambiguity of evaluation point in a stochastic integral. However, the underlying SDEs and entropy are not, and should not be, ambiguous since we have specified Ito SDEs and have used the short time propagator appropriate for their corresponding Fokker-Planck equation. Since all evaluation points lead to the correct path probability density we are not obliged to consider, for example, only Ito-type multiplication ($a = b = 0$) simply because the underlying SDEs are of Ito form. Rather, to proceed we recognise that with multiplicative noise we must ensure that we evaluate the two transition probability densities at precisely the same coordinates and not just at same time (which would suffice for additive noise), for the same reasons that make stochastic integration sensitive to the specific integration scheme when the integrand has dependence on the integrating variable. Alternatively it may be reasoned that as $dt \rightarrow 0$ we require the short time propagators to approach jump transition probabilities of a master equation. Under such a description the entropy production can be unambiguously described by the ratios of probabilities appearing in a master equation approach [19, 23]. Therefore, if we represent such a quantity using the short time propagators we require the transition rates in both numerator and denominator to be evaluated equivalently. Since these are characterised by our system variables A_i, B_i etc, to effect such a condition we require $\mathbf{r}' = \mathbf{r}$ which is equivalent to making the choice $b = 1 - a$, noting that a is still a free parameter. One may think of this as insisting that the path transformation $\vec{x}^\dagger(t) = \varepsilon \vec{x}(\tau - t)$ persists on a (sub) infinitesimal scale, which would not matter in normal calculus, so that the noise is experienced in precisely the right way. Inserting $b = 1 - a$ into the above yields Eq. (28) which has no dependence on the choice a , an indication that it is

defined in a sound fashion.

Appendix B: Consistency of entropy contributions

By construction, we have $\Delta S_{\text{tot}} = \Delta S_1 + \Delta S_2 + \Delta S_3$ which according to the expressions in Eqs. (37), (53), (61) and (67) means we require

$$\begin{aligned} \frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = & \sum_i \int d\mathbf{x} \frac{(J_i^{\text{ir}}(\mathbf{x}))^2}{p(\mathbf{x}) D_i(\mathbf{x})} \\ = & \sum_i \int d\mathbf{x} \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i^{\text{ir}}(\mathbf{x})}{p(\mathbf{x})} - \frac{J_i^{\text{st,ir}}(\mathbf{x})}{p^{\text{st}}(\mathbf{x})} \right)^2 \\ & + \int d\mathbf{x} \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i^{\text{ir,st}}(\varepsilon \mathbf{x})}{p^{\text{st}}(\varepsilon \mathbf{x})} \right)^2 \\ & + \int d\mathbf{x} J_i(\mathbf{x}) (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon \mathbf{x})). \end{aligned} \quad (\text{B1})$$

Thus in order that everything is consistent we require

$$\begin{aligned} 0 = & \sum_i \int d\mathbf{x} \left[-2 \frac{J_i^{\text{ir}}(\mathbf{x})}{D_i(\mathbf{x})} \frac{J_i^{\text{ir,st}}(\mathbf{x})}{p^{\text{st}}(\mathbf{x})} + \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i^{\text{ir,st}}(\mathbf{x})}{p^{\text{st}}(\mathbf{x})} \right)^2 \right. \\ & \left. + \frac{p(\mathbf{x})}{D_i(\mathbf{x})} \left(\frac{J_i^{\text{ir,st}}(\varepsilon \mathbf{x})}{p^{\text{st}}(\varepsilon \mathbf{x})} \right)^2 + J_i(\mathbf{x}) (\phi'_i(\mathbf{x}) - \varepsilon_i \phi'_i(\varepsilon \mathbf{x})) \right]. \end{aligned} \quad (\text{B2})$$

By substitution of the definitions of the fluxes this reduces to

$$\begin{aligned} 0 = & \sum_i \int d\mathbf{x} \left[2A_i^{\text{ir}}(\mathbf{x}) \frac{\partial p(\mathbf{x})}{\partial x_i} - 2 \frac{\partial D_i(\mathbf{x})}{\partial x_i} \frac{\partial p(\mathbf{x})}{\partial x_i} \right. \\ & + \frac{\partial(D_i(\mathbf{x})p(\mathbf{x}))}{\partial x_i} (\phi'_i(\mathbf{x}) + \varepsilon_i \phi'_i(\varepsilon \mathbf{x})) \\ & + p(\mathbf{x}) D_i(\mathbf{x}) ((\phi'_i(\mathbf{x}))^2 + (\phi'_i(\varepsilon \mathbf{x}))^2) \\ & - 2p(\mathbf{x}) \frac{\partial D_i(\mathbf{x})}{\partial x_i} (\phi'_i(\mathbf{x}) + \varepsilon_i \phi'_i(\varepsilon \mathbf{x})) \\ & + p(\mathbf{x})(A_i^{\text{ir}}(\mathbf{x}) + A_i^{\text{rev}}(\mathbf{x})) \phi'_i(\mathbf{x}) \\ & \left. + \varepsilon_i p(\mathbf{x})(A_i^{\text{ir}}(\mathbf{x}) - A_i^{\text{rev}}(\mathbf{x})) \phi'_i(\varepsilon \mathbf{x}) \right]. \end{aligned} \quad (\text{B3})$$

Integrating by parts, dropping or cancelling boundary terms and using the definition of the irreversible and reversible drift terms yields the condition

$$\begin{aligned} 0 = & \sum_i \int d\mathbf{x} p(\mathbf{x}) \left[-2 \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} + 2 \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} \right. \\ & - D_i(\mathbf{x}) (\phi''_i(\mathbf{x}) + \phi''_i(\varepsilon \mathbf{x})) + D_i(\mathbf{x}) ((\phi'_i(\mathbf{x}))^2 + (\phi'_i(\varepsilon \mathbf{x}))^2) \\ & - 2 \frac{\partial D_i(\mathbf{x})}{\partial x_i} (\phi'_i(\mathbf{x}) + \varepsilon_i \phi'_i(\varepsilon \mathbf{x})) + (A_i^{\text{ir}}(\mathbf{x}) + A_i^{\text{rev}}(\mathbf{x})) \phi'_i(\mathbf{x}) \\ & \left. + (A_i^{\text{ir}}(\varepsilon \mathbf{x}) + A_i^{\text{rev}}(\varepsilon \mathbf{x})) \phi'_i(\varepsilon \mathbf{x}) \right]. \end{aligned} \quad (\text{B4})$$

The divergenceless stationary distribution condition however yields

$$0 = \left(- (A_i^{\text{ir}}(\mathbf{x}) + A_i^{\text{rev}}(\mathbf{x}))\phi'_i(\mathbf{x}) - D_i(\mathbf{x})(\phi'_i(\mathbf{x}))^2 + \frac{\partial A_i^{\text{ir}}(\mathbf{x})}{\partial x_i} + \frac{\partial A_i^{\text{rev}}(\mathbf{x})}{\partial x_i} - \frac{\partial^2 D_i(\mathbf{x})}{\partial x_i^2} + D_i(\mathbf{x})\phi''_i(\mathbf{x}) + 2\frac{\partial D_i(\mathbf{x})}{\partial x_i}\phi'_i(\mathbf{x}) \right) e^{-\phi(\mathbf{x})}, \quad (\text{B5})$$

but also

$$0 = \left(- (A_i^{\text{ir}}(\varepsilon\mathbf{x}) + A_i^{\text{rev}}(\varepsilon\mathbf{x}))\phi'_i(\varepsilon\mathbf{x}) - D_i(\varepsilon\mathbf{x})(\phi'_i(\varepsilon\mathbf{x}))^2 + \frac{\partial A_i^{\text{ir}}(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)} + \frac{\partial A_i^{\text{rev}}(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)} - \frac{\partial^2 D_i(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)^2} + D_i(\varepsilon\mathbf{x})\phi''_i(\varepsilon\mathbf{x}) + 2\frac{\partial D_i(\varepsilon\mathbf{x})}{\partial(\varepsilon_i x_i)}\phi'_i(\varepsilon\mathbf{x}) \right) e^{-\phi(\varepsilon\mathbf{x})}. \quad (\text{B6})$$

Combining the two conditions in Eqs. (B5) and (B6) yields the contents of the brackets in equation (B4) and so the result is proved.

[1] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).

[2] D. J. Evans and D. J. Searles, Phys. Rev. E **52**, 5839 (1995).

[3] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).

[4] D. Carberry, J. C. Reid, G.M.Wang, E. Sevick, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **92**, 140601 (2004).

[5] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).

[6] J. Kurchan, J. Phys. A: Math. Gen. **31**, 3719 (1998).

[7] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).

[8] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[9] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[10] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[11] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[12] K. Sekimoto, Prog. Theor. Phys. Suppl. **130**, 17 (1998).

[13] T. Hatano and S. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).

[14] T. Speck and U. Seifert, J. Phys. A: Math. Gen. **38**, L581 (2005).

[15] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, J. Stat. Mech. P08001 (2006).

[16] M. Esposito, U. Harbola, and S. Mukamel, Phys. Rev. E **76**, 031132 (2007).

[17] H. Ge, Phys. Rev. E **80**, 021137 (2009).

[18] H. Ge and H. Qian, Phys. Rev. E **81**, 051133 (2010).

[19] M. Esposito and C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010).

[20] M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010).

[21] C. Van den Broeck and M. Esposito, Phys. Rev. E **82**, 011144 (2010).

[22] Y. Oono and M. Paniconi, Prog. Theor. Phys. Suppl. **130**, 29 (1998).

[23] R. E. Spinney and I. J. Ford, arXiv:1201.0904.

[24] R. J. Harris and G. M. Schütz, J. Stat. Mech. P07020 (2007).

[25] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).

[26] C. Wissel, Z. Physik B **35**, 185 (1979).

[27] T. Tomé and M. J. de Oliveira, Phys. Rev. E **82**, 021120 (2010).

[28] D. Daems and G. Nicolis, Phys. Rev. E **59**, 4000 (1999).

[29] M. Matsuo and S. Sasa, Physica A **276**, 188 (2000).

[30] N. V. Kampen, J. Phys. Chem. Sol. **49**, 673 (1988).

[31] G. Stolovitzky, Phys. Lett. A **241**, 240 (1998).

[32] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics, Vol. 799 (Springer, Berlin Heidelberg, 2010).

[33] M. Widder and U. Titulaer, Physica A **154**, 452 (1989).

[34] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995).

[35] K. Sekimoto, Prog. Theor. Phys. Suppl. **130**, 17 (1998).

[36] A. Saha, S. Lahiri, and A. M. Jayannavar, Phys. Rev. E **80**, 011117 (2009).

[37] B. H. Shargel, J. Phys. A: Math. Gen. **43**, 135002 (2010).

[38] T. Speck, V. Bickle, C. Bechinger, and U. Seifert, Europhys. Lett. **79**, 30002 (2007).